Financial Frictions and the Wealth Distribution

Jesús Fernández-Villaverde

University of Pennsylvania, NBER, and CEPR

Samuel Hurtado — Banco de España

Galo Nuño — Banco de España*

March 2, 2020

Abstract

We postulate a nonlinear DSGE model with a financial sector and heterogeneous households. In our model, the interaction between the supply of bonds by the financial sector and the precautionary demand for bonds by households produces significant *endogenous aggregate risk*. This risk induces an endogenous regime-switching process for output, the risk-free rate, excess returns, debt, and leverage. The regime-switching generates i) multimodal distributions of the variables above; ii) time-varying levels of volatility and skewness for the same variables; and iii) supercycles of borrowing and deleveraging. All of these are important properties of the data. In comparison, the representative household version of the model cannot generate any of these features. Methodologically, we discuss how nonlinear DSGE models with heterogeneous agents can be efficiently computed using machine learning and how they can be estimated with a likelihood function, using inference with diffusions.

Keywords: Heterogeneous agents; wealth distribution; financial frictions; continuous-time; machine learning; neural networks; structural estimation; likelihood function.

JEL codes: C45, C63, E32, E44, G01, G11.

*A `GitHub` repository with all the codes for the paper and further examples is at https://github.com/jesusfv/financial-frictions. We thank Manuel Arellano, Emmanuel Farhi, Xavier Gabaix, Lars Peter Hansen, Mark Gertler, Aubhik Khan, Davide Melcangi, Ben Moll, Chris Sims, Gianluca Violante, Ivan Werning, and participants at numerous seminars and conferences for pointed comments. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the Eurosystem or the Bank of Spain.

1 Introduction

Recently, several researchers have documented the nonlinear relations between financial variables and aggregate fluctuations. For example, Jordà et al. (2016) have gathered data from 17 advanced economies over 150 years to show how output growth, volatility, skewness, and tail events all seem to depend on the levels of leverage in an economy. Similarly, Adrian et al. (2019b) have found how, in the U.S., sharply negative output growth follows worsening financial conditions associated with leverage.

Consequently, we need to postulate fully nonlinear dynamic stochastic general equilibrium (DSGE) models with a financial sector that can account for the empirical observations in these papers. After suitable development, such DSGE models will serve as laboratories for policy analysis and forecasting.

This paper takes a step in this direction by building, computing, and estimating a continuous-time DSGE economy with a financial sector, modeled as a representative financial expert, and households, subject to uninsurable idiosyncratic labor productivity shocks. The economy is hit by aggregate shocks to the stock of capital, and it is subject to two financial frictions. First, only the expert can hold the capital that is rented to a representative firm, and it cannot issue state-contingent assets (i.e., outside equity) to finance it. Instead, the expert can issue a risk-free bond to leverage its equity and accumulate more capital. Hence, the expert must absorb all the capital-return risk. Second, the households can only save in the risk-free bond to self-insure against idiosyncratic shocks and the variations in wages and the risk-free rate induced by aggregate fluctuations. The bond, in addition, cannot be shorted.

We include a financial sector in our model to capture the evolution of debt and leverage. We introduce household heterogeneity because, under the financial frictions we consider, such a heterogeneity begets substantial *endogenous aggregate risk*. This aggregate risk derives from the interaction between the supply of bonds by the expert, determined by the excess return of capital over risk-free bonds, and the precautionary demand for bonds by the households, driven by their income risk.

The endogenous aggregate risk induces an *endogenous regime-switching process* for output, the risk-free rate, excess returns, debt, and leverage. The regime-switching generates i) multimodal distributions of the variables above (as reported in Adrian et al., 2019a); ii) time-varying levels of volatility and skewness for the same variables (as documented in Fernández-Villaverde and Guerrón-Quintana, 2020); and iii) supercycles of borrowing and deleveraging (as evidenced in Reinhart and Rogoff, 2009).

Thus, household heterogeneity allows us not only to replicate the nonlinear dynamics between aggregate variables and leverage reported by Jordà et al. (2016) and Adrian et al. (2019b), but also to offer a novel and simultaneous account of i) the recent heightened fragility of the advanced economies to adverse shocks; ii) the rise in wealth inequality witnessed before the 2007-2009

financial crisis (Alvaredo et al., 2017); iii) the increase in debt and leverage experienced during the same period (Adrian and Shin, 2010, and Nuño and Thomas, 2017); and iv) the low risk-free interest rates of the last two decades (Holston et al., 2017).

Let us unpack the previous three paragraphs. For parameter values that match U.S. data and maximize the likelihood function, our economy has multiple stochastic steady states, or SSS(s), despite the model having a unique deterministic steady state (DSS).[1] In particular, we will have a high-leverage SSS (HL-SSS) and a low-leverage SSS (LL-SSS), each with its basin of attraction (there is a third, unstable SSS that we do not need to discuss).

The intuition for the existence of two stable SSS(s) is as follows. In the basin of attraction of the HL-SSS, endogenous aggregate risk is high. After a negative aggregate shock, the economy suffers deep and protracted recessions with persistently low wages (the next paragraph explains why). Due to precautionary behavior, households accumulate more bonds when risk is high. Higher savings have two consequences. First, higher savings increase wealth inequality among households: households with good realizations of their idiosyncratic shock accumulate much wealth. Second, higher savings lower the risk-free rate that clears the bond market and cause a high expected excess return for capital, pushing the expert to leverage aggressively, the defining feature of this basin of attraction. In comparison, aggregate risk is low in the basin of attraction of the LL-SSS. Therefore, precautionary savings are lower, wealth inequality smaller, the risk-free rate high, and the expected excess return reduced. These prices sustain the low leverage that characterizes this basin.

Why are recessions more severe in the basin of attraction of the HL-SSS than around the LL-SSS? When leverage is high, a negative aggregate shock to capital greatly erodes the expert's net wealth, since its relatively small equity must absorb all the capital losses. After a negative shock, it takes a long time for the expert's wealth and capital to recover. During this time, wages are persistently low and the risk-free rate is high.[2] Two forces cause the sluggish recovery of capital. First, the supply of debt is low because the expert is reluctant to lever too quickly. Second, the demand for debt is low because wealthy households, whose income is more dependent on bond returns, have a weaker incentive to reduce their consumption when the risk-free rate is persistently high. These two forces result in small debt growth and the slow recovery of capital and output. By contrast, when leverage is low, the recessions after a negative aggregate shock are mild. While the responses are similar on impact, the ensuing recovery is much faster because the expert experiences less of a reduction in its net wealth.

Whether the economy has high or low leverage is a consequence of past aggregate shocks.

[1] An SSS, also known as a risky steady state, is a fixed point of the equilibrium conditions of the model when the *realization* of the aggregate shock is zero. A DSS is a fixed point of the equilibrium conditions of the model when the *volatility* of the aggregate shock (but not of idiosyncratic shocks) is zero. The multiplicity of SSS(s) is different from the multiplicity of equilibria: in our model, we find a *unique* equilibrium.

[2] In our model, wages are equal to the marginal productivity of labor (lower when capital is low) and the risk-free interest rate depends on the marginal productivity of capital (higher when capital is low).

Sometimes, while the economy is traveling in the basin of attraction of the HL-SSS, a sequence of aggregate shocks will move it to the basin of attraction of the LL-SSS (and vice versa). For our baseline parameter values, the economy will spend more time, on average, in the basin of attraction of the HL-SSS than in the basin of attraction of the LL-SSS. This finding, however, varies as we change the volatility of idiosyncratic and aggregate shocks.

The transition between the two basins of attraction amounts to an *endogenous regime-switching* process for output, interest rates, debt, and leverage and a strong state-dependence in the impulse response functions of the model. This endogenous switching generates a bimodal and skewed ergodic distribution of aggregate variables, time-variation in the volatility and skewness of these variables, and supercycles of borrowing and deleveraging. Interestingly, these supercycles may last for decades, with their associated long-lasting changes in aggregate volatility and wealth inequality. Consider, for instance, the supercycle of the advanced economies between 1945 and nowadays, as leverage was slowly rebuilt after World War II, pushing the economy toward higher levels of debt, wealth inequality, and financial fragility (we will discuss a complementary channel through higher micro turbulence below). Our model offers a parsimonious interpretation of such a supercycle and its duration.

Our findings are in stark contrast to the properties of the model when we move it toward its representative household version. As we reduce idiosyncratic risk (while keeping the volatility of the aggregate shock constant), the precautionary saving motive becomes smaller and the excess return is low. Eventually, the model encounters a bifurcation and the HL-SSS evaporates. Conversely, when we increase idiosyncratic risk, the LL-SSS vanishes. In this case, households are so concerned about their idiosyncratic risk that their demand for bonds propels the risk-free rate sufficiently low to allow only the existence of the HL-SSS (paradoxically, increasing the endogenous aggregate risk they face). Our findings also diverge from the version of the model where the expert can issue outside equity without constraints. In such a case, the expert and the households share the capital risk, and the large impact of leverage movements (and its associated nonlinearities) disappears.

Since household heterogeneity is at the core of our argument, changes in the forces behind precautionary savings affect the overall behavior of the economy. For example, an increase in idiosyncratic risk, such as the one documented for the U.S. since the late 1970s by Moffitt and Zhang (2018), translates in our model into higher macro volatility –even when the variance of aggregate shocks remains constant– and the four empirical phenomena highlighted above: i) higher wealth inequality, ii) higher debt and leverage, iii) low risk-free interest rates, and iv) heightened fragility of the economy to adverse shocks.[3]

This connection between micro and macro volatility is different from the "paradox of volatil-

[3]In the data, the economy is buffeted by many aggregate shocks. Some of them, such as monetary policy shocks, have become less volatile since the 1980s, thus lowering the overall volatility of the economy despite higher micro turbulence. Our claim is a comparative statics statement *with respect to* capital shocks.

ity" in Brunnermeier and Sannikov (2014). In their model, a low volatility of aggregate shocks leads to higher leverage by the financial expert and, thus, deep recessions when a large shock hits the economy. In our model, a high volatility of idiosyncratic shocks leads to higher leverage and deeper recessions despite the volatility of aggregate shocks being constant.[4]

Solving and structurally estimating our economy is daunting. Since the wealth distribution is an infinite-dimensional object, standard dynamic programming techniques cannot be employed. To meet this challenge, our paper provides new tools for the global, nonlinear solution and structural estimation of heterogeneous agent models with aggregate shocks.

In terms of global, nonlinear solution methods, we rely on machine learning and employ a neural network to obtain a flexible, nonlinear approximation of the perceived law of motion (PLM) of the cross-sectional distribution of assets (the expert's equity and households' bonds) with a finite set of moments.[5]

On the one hand, we can approximate the PLM with high accuracy using moments of the distribution à la Krusell and Smith (1998) because the consumption decision rule of the households is close to linear with respect to the household state variables (except for impoverished households). On the other hand, we need the PLM to be a general function of the moments of the distribution because the consumption decision rule of the households is sharply nonlinear with respect to the aggregate state variables (equity and debt). Households understand the strongly state-dependent responses of wages and the risk-free rate with respect to the aggregate amounts of debt and equity and act accordingly. Neural networks can capture a nonlinear PLM without having to specify, ex-ante, any concrete structure for it. All linear models are alike, but every nonlinear model is nonlinear in its particular and unpredictable way.

In comparison, in Krusell and Smith (1998), the PLM of the aggregate variables is loglinear in the endogenous aggregate state variables.[6] This PLM is a poor choice in our model because of nonlinearities created by the significant level of endogenous risk in the economy. To show this, we report how a naive implementation of the Krusell and Smith (1998) algorithm in our economy, and even of refinements such as Chebyshev polynomials, delivers much worse accuracy.

While many nonlinear solution schemes are possible, our machine learning approach is convenient, both regarding theoretical properties and practical considerations. First, the *universal approximation theorem* (Hornik et al., 1989; Cybenko, 1989; Bach, 2017) states that a neural network can approximate any unknown Borel measurable function. Second, the neural network breaks the curse of dimensionality for a large class of approximated functions, which allows our

[4]Our mechanism is also different from that in Kumhof et al. (2015). In their model, the change in the wealth distribution is due to an exogenous shock to the top earners' income. In our model, the change is endogenous.

[5]An interactive `GitHub` repository with all the codes for the paper and further examples of our methodology is available at https://github.com/jesusfv/financial-frictions.

[6]In Krusell and Smith (1998), the PLM is nonlinear in the exogenous states because the coefficients of the regression are allowed to vary across shocks. However, in our model, exogenous states are incorporated into the endogenous states instantaneously and this refinement is moot.

method to be extended to environments with many state variables (or to use, if needed, rich features of the distribution such as fine quantiles). Third, the neural network can be efficiently trained using a combination of the *gradient descent* and the *back-propagation* algorithms. Fourth, our algorithm is easy to code and readily amenable to massive parallelization in GPUs, FPGAs, and dedicated AI accelerators such as TPUs. Fifth, our approach transparently reflects the self-justified equilibria nature of the "bounded rationality" solution of most heterogeneous agents models (Kubler and Scheidegger, 2018). The PLM is computed based on the samples drawn in the simulation of paths within the aggregate ergodic distribution. The agents employ the neural network to extrapolate the dynamics outside of the equilibrium region.

Continuous time helps us to characterize much of the equilibrium dynamics analytically and to avoid dealing with expectations even when solving the model globally. However, nothing essential depends on this choice. We could replicate our approach –with higher computational costs– in discrete time. Achdou et al. (2017) and Nuño and Thomas (2016) provide a general discussion of the advantages of continuous-time methods.

In terms of structural estimation, we illustrate how a fully nonlinear model can be taken to the data with a likelihood function employing aggregate and micro observations using inference with diffusions (Lo, 1988). The likelihood is computationally straightforward once we have solved the model with the approach outlined above: it just amounts to transposing a matrix. Likelihood-based estimation is attractive in this context because methods of moments (either informal, such as calibration, or formal, such as the SMM) face difficulties when dealing with multimodal ergodic distributions of the moments of interest (Andreasen et al., 2018).

Our work is related to important threads in macroeconomics. First, we follow the macro-finance literature pioneered by Basak and Cuoco (1998), Adrian and Boyarchenko (2012), He and Krishnamurthy (2012, 2013), and Brunnermeier and Sannikov (2014), among others. Most of these papers only consider *between-agents* heterogeneity (e.g., a representative household vs. a representative expert), but no *within-agents* heterogeneity (e.g., different households), the key to our results. A paper with *within-agents* heterogeneity, but without aggregate shocks, is Biggio and Sannikov (2019).

Our paper also contributes to the literature on global solution methods for heterogeneous agents models with aggregate shocks such as Den Haan (1996, 1997), Algan et al. (2008), Reiter (2009, 2010), Den Haan and Rendahl (2010), Maliar et al. (2010), Sager (2014), Pröhl (2015), Bayer and Luetticke (2018), Winberry (2018) (who also proposes an estimation algorithm), and Auclert et al. (2019) (Algan et al., 2014, is a survey of the field). To the best of our knowledge, we are the first to generalize the celebrated algorithm of Krusell and Smith (1998) to accommodate a universal nonlinear law of motion in the endogenous state variables.[7]

[7] Ahn et al. (2017) introduce a related method to compute the solution to heterogeneous agents models with aggregate shocks in continuous time. However, theirs is a local solution, based on first-order perturbation around the DSS and, thus, unable to analyze the nonlinear dynamics posed by our paper.

Finally, our paper builds on the nascent literature on the application of machine learning techniques to compute dynamic models. The proposed methods have so far been concerned with the solution of high-dimensional dynamic programming problems. Examples include Scheidegger and Bilionis (2017), Duarte (2018), Maliar et al. (2019), and Azinović et al. (2020). Instead, our machine learning algorithm is used to provide a nonlinear forecast of aggregate variables within the model. In this respect, our paper reconnects with an early literature using neural networks to model bounded rationality and learning, such as Cho (1995) and Cho and Sargent (1996).

2 Model

We postulate a continuous-time, infinite-horizon DSGE model with three types of agents: a representative firm, a representative financial expert, and a continuum of households. There is one risky asset, capital, and a risk-free one, noncontingent bonds. Only the expert can hold the risky asset. In the interpretation implicit in our terminology, this is because the expert is the only agent with knowledge in accumulating capital. However, other interpretations, such as the expert standing in for the financial sector, are possible. In contrast, households can lend to the expert at the risk-free rate, but cannot hold capital themselves, as they lack the required skill to handle it. The expert cannot issue outside equity, but it can partially finance its holdings of the risky asset by issuing bonds to households. Together with market clearing, our assumptions imply that, at the aggregate level, there is a positive net supply of capital, while bonds are in zero net supply. As will become apparent below, there is no need to distinguish between the firm and the expert, and we could write the model consolidating both agents into a single type. Keeping both agents separate, though, clarifies the exposition. We introduce heterogeneity on the side of the households –but not among the financial experts or the firms– because this heterogeneity triggers significant *endogenous aggregate risk*, the key feature of our investigation.

2.1 The firm

A representative firm rents aggregate capital, K_t, and aggregate labor, L_t, to produce output with a Cobb-Douglas technology $Y_t = F(K_t, L_t) = K_t^{\alpha} L_t^{1-\alpha}$. Since input markets are competitive, wages, w_t, are equal to the marginal productivity of labor:

$$w_t = \frac{\partial F\left(K_t, L_t\right)}{\partial L_t} = (1-\alpha)\frac{Y_t}{L_t} \tag{1}$$

and the rental rate of capital, rc_t, is equal to the marginal productivity of capital:

$$rc_t = \frac{\partial F\left(K_t, L_t\right)}{\partial K_t} = \alpha \frac{Y_t}{K_t}. \tag{2}$$

During production, capital depreciates at a constant rate δ and receives a growth rate shock Z_t that follows a Brownian motion with volatility σ. Thus, aggregate capital evolves as:

$$\frac{dK_t}{K_t} = (\iota_t - \delta)\, dt + \sigma dZ_t, \tag{3}$$

where ι_t is the reinvestment rate per unit of capital that we will characterize below. The growth rate shock Z_t is a common assumption in the financial frictions literature because it impacts the equity of the financial expert directly, making the effects of the shock transparent (for example, see Brunnermeier and Sannikov, 2014). The shock can be understood as a stand-in for any type of exogenously given changes in the efficiency units of capital. For example, it can represent the consequences of a new environmental regulation that lowers the profitability of existing capital or a technological innovation that renders already-installed capital economically obsolete (see Greenwood et al., 1997, for details).

We define the rental rate of capital rc_t over the capital contracted, K_t, and not over the capital returned after depreciation and the growth rate shock. Thus, the instantaneous return rate on capital is $dr_t^k = (rc_t - \delta)\, dt + \sigma dZ_t$. The coefficient of the time drift, $rc_t - \delta$, is the profit rate of capital, equal to the rental rate of capital less depreciation. The volatility σ determines the capital gains rate.

2.2 The expert

The representative financial expert holds capital $\widehat{K}_t$ (we denote variables related to the expert with a caret). It rents this capital to the firm. To finance its holding of $\widehat{K}_t$, the expert issues risk-free debt $\widehat{B}_t$ at rate r_t to the households. The financial frictions in the model come from the fact that the expert cannot issue state-contingent claims (i.e., outside equity) against $\widehat{K}_t$. In particular, the expert must absorb all the risk from holding capital.

The net wealth (i.e., inside equity) of the expert, $\widehat{N}_t$, is the difference between its assets (capital) and liabilities (debt), $\widehat{N}_t = \widehat{K}_t - \widehat{B}_t$. We allow $\widehat{N}_t$ to be negative, although this will not occur along the equilibrium path.

Let $\widehat{C}_t$ be the consumption of the expert. Then, $\widehat{N}_t$ evolves as:

$$\begin{aligned} d\widehat{N}_t &= \widehat{K}_t dr_t^k - \widehat{B}_t r_t dt - \widehat{C}_t dt \\ &= \left[\left(r_t + \widehat{\omega}_t \left(rc_t - \delta - r_t\right)\right) \widehat{N}_t - \widehat{C}_t\right] dt + \sigma \widehat{\omega}_t \widehat{N}_t dZ_t, \end{aligned} \tag{4}$$

where $\widehat{\omega}_t \equiv \frac{\widehat{K}_t}{\widehat{N}_t}$ is the leverage ratio of the expert. The term $r_t + \widehat{\omega}_t \left(rc_t - \delta - r_t\right)$ is the deterministic return on net wealth, equal to the return on bonds, r_t, plus $\widehat{\omega}_t$ times the excess return on leverage, $rc_t - \delta - r_t$. The term $\sigma \widehat{\omega}_t \widehat{N}_t$ reflects the risk of holding capital induced by the capital growth

rate shock. Equation (4) allows us to derive the law of motion for $\widehat{K}_t$:

$$d\widehat{K}_t = d\widehat{N}_t + d\widehat{B}_t = \left[\left(r_t + \widehat{\omega}_t \left(rc_t - \delta - r_t\right)\right)\widehat{N}_t - \widehat{C}_t\right] dt + \sigma\widehat{\omega}_t\widehat{N}_t dZ_t + d\widehat{B}_t.$$

The expert's preferences over $\widehat{C}_t$ can be represented by:

$$\widehat{U}_j = \mathbb{E}_j\left[\int_j^\infty e^{-\widehat{\rho}(t-j)}\log(\widehat{C}_t)dt\right], \tag{5}$$

where $\widehat{\rho}$ is its discount rate. The log felicity function will make our derivations below easier, but it could be easily generalized to recursive preferences à la Duffie and Epstein (1992).

The expert decides its consumption levels and leverage ratio to solve the problem:

$$\max_{\{\widehat{C}_t,\widehat{\omega}_t\}_{t\geq 0}} \widehat{U}_0, \tag{6}$$

subject to the evolution of N_t (4), an initial net wealth N_0, and the no-Ponzi-game condition:

$$\lim_{T\to\infty} e^{-\int_0^T r_\tau d\tau} B_T = 0. \tag{7}$$

2.3 Households

There is a continuum of infinitely lived households with unit mass. Households are heterogeneous in their wealth a_m and labor supply z_m for $m \in [0,1]$. The distribution of households at time t over these two individual states is $G_t(a,z)$. To save on notation, we will drop the subindex m when no ambiguity occurs.

Each household supplies z_t units of labor valued at wage w_t. Idiosyncratic labor productivity evolves stochastically following a two-state Markov chain: $z_t \in \{z_1, z_2\}$, with $0 < z_1 < z_2$. The process jumps from state 1 to state 2 with intensity λ_1 and vice versa with intensity λ_2. The ergodic mean of z is 1. As in Huggett (1993), we identify state 1 with unemployment (where z_1 is the value of leisure and home production) and state 2 with working. We will follow this assumption when the model faces the data, but nothing essential depends on it. Also, increasing the number of states of the chain is trivial, but two points will suffice for our purposes.

Households can save an amount a_t in the riskless debt issued by the expert at interest rate r_t. Hence, a household's wealth follows:

$$da_t = \left(w_t z_t + r_t a_t - c_t\right) dt = s\left(a_t, z_t, K_t, G_t\right) dt, \tag{8}$$

where $s(a_t, z_t, K_t, G_t)$ denotes the drift of the wealth process. The first two variables, a_t and z_t, are the household individual states, and the next two, K_t and G_t, are the aggregate state

variables that determine the returns on its income sources (labor and bonds). All four variables pin down the optimal choice, $c_t = c(a_t, z_t, K_t, G_t)$, of the control. The households also face a borrowing limit that prevents them from shorting bonds:

$$a_t \geq 0. \tag{9}$$

Households have a CRRA instantaneous felicity function $u(c_t) = \frac{c_t^{1-\gamma}-1}{1-\gamma}$ discounted at rate $\rho > 0$. We pick this functional form to allow different risk aversions in the households and the expert, but we could substitute it with more general recursive preferences.

Also, we make the households less patient than the expert, $\rho > \widehat{\rho}$. We will show later how the risk-free rate in the DSS (recall, the deterministic steady state) is pinned down by the discount factor of the expert, i.e., $r = \widehat{\rho}$ (we drop the subindex when we denote a variable evaluated at the DSS). But if $\rho \leq r = \widehat{\rho}$, the households would want to accumulate savings without bounds to self-insure against idiosyncratic labor risk (Aiyagari, 1994). Hence, we can only have a DSS –and an associated ergodic distribution of individual endogenous variables– if we increase the households' discount rate above the expert's.[8]

In summary, households maximize

$$\max_{\{c_t\}_{t\geq 0}} \mathbb{E}_0 \left[\int_0^\infty e^{-\rho t} \frac{c_t^{1-\gamma}-1}{1-\gamma} dt \right], \tag{10}$$

subject to the budget constraint (8), initial wealth a_0, and the borrowing limit (9).

2.4 Market clearing

There are three market clearing conditions. First, the total amount of debt issued by the expert must equal the total amount of households' savings:

$$B_t \equiv \int a dG_t(a, z) = \widehat{B}_t, \tag{11}$$

which also implies $dB_t = d\widehat{B}_t$.

Second, the total amount of labor rented by the firm is equal to labor supplied, $L_t = \int z dG_t$. Since the ergodic mean of z is 1, we have that $L_t = 1$ and total payments to labor are w_t.

[8] This property of our economy stands in contrast with models à la Bernanke et al. (1999), where borrowers are more impatient than lenders to prevent the former from accumulating enough wealth as to render the financial friction inoperative. However, in these models, borrowers are infinitesimal and subject to idiosyncratic risk, and the lenders' discount rate determines the DSS risk-free rate. The situation is reversed in our model, with the lenders being infinitesimal and subject to idiosyncratic risk and the borrower's discount rate controlling the DSS risk-free rate. We have framed our discussion for the case without aggregate shocks, since we want to ensure the existence of a DSS. The characterization of the admissible region for ρ in relation to $\widehat{\rho}$ when we only care about the properties of the economy with aggregate shocks is beyond the scope of our paper.

If we define total consumption by households as

$$C_t \equiv \int c\left(a_t, z_t, K_t, G_t\right) dG_t\left(a, z\right),$$

we get:

$$d\widehat{B}_t = dB_t = \left(w_t + r_t B_t - C_t\right) dt, \tag{12}$$

which tells us that the evolution of aggregate debt is the labor income of households (w_t) plus its debt income ($r_t B_t$) minus their aggregate consumption C_t.

Third, the total amount of capital in this economy is owned by the expert, $K_t = \widehat{K}_t$, and, therefore, $dK_t = d\widehat{K}_t$ and $\widehat{\omega}_t = \frac{K_t}{N_t}$, where $N_t = \widehat{N}_t = K_t - B_t$. With these results, we derive

$$\begin{aligned} dK_t &= \left(\left(r_t + \widehat{\omega}_t\left(rc_t - \delta - r_t\right)\right)\widehat{N}_t - \widehat{C}_t\right) dt + \sigma\widehat{\omega}_t\widehat{N}_t dZ_t + d\widehat{B}_t \\ &= \left(Y_t - \delta K_t - C_t - \widehat{C}_t\right) dt + \sigma K_t dZ_t, \end{aligned} \tag{13}$$

where the last line uses the fact that, from competitive input markets and constant-returns-to-scale, $Y_t = rc_t K_t + w_t$. Recall, from equation (3), that $dK_t = \left(\iota_t - \delta\right) K_t dt + \sigma K_t dZ_t$. Then, equating (13) and (3) and cancelling terms, we get

$$\iota_t = \frac{Y_t - C_t - \widehat{C}_t}{K_t},$$

i.e., the reinvestment rate is output less aggregate consumption divided by aggregate capital.

2.5 Density

The households' distribution $G_t\left(a, z\right)$ has a density on assets a, $g_{it}(a)$, conditional on the labor productivity state $i \in \{1, 2\}$. The density satisfies the normalization

$$\sum_{i=1}^{2} \int_0^\infty g_{it}(a) da = 1.$$

The dynamics of this density conditional on the realization of aggregate variables are given by the Kolmogorov forward (KF) equation:

$$\frac{\partial g_{it}}{\partial t} = -\frac{\partial}{\partial a}\left(s\left(a, z, K_t, G_t\right) g_{it}(a)\right) - \lambda_i g_{it}(a) + \lambda_j g_{jt}(a), \ i \neq j = 1, 2. \tag{14}$$

Reading equation (14) is simple: the density evolves according to the optimal consumption-saving choices of each household plus two jumps corresponding to households that circulate out of the labor state i ($\lambda_i g_{it}(a)$) and the households that move into state j ($\lambda_j g_{jt}(a)$).

3 Equilibrium

An equilibrium in this economy is composed of a set of prices $\left\{w_t, rc_t, r_t, r_t^k\right\}_{t\geq 0}$, quantities $\left\{K_t, N_t, B_t, \widehat{C}_t, c_{mt}\right\}_{t\geq 0}$ and a density $\left\{g_{it}\left(\cdot\right)\right\}_{t\geq 0}$ for $i \in \{1,2\}$ such that:

1. Given w_t, r_t, and g_t, the solution of household m's problem (10) is $c_{mt} = c\left(a_t, z_t, K_t, G_t\right)$.
2. Given r_t^k, r_t, and N_t, the solution of the expert's problem (6) is $\widehat{C}_t$, K_t, and B_t.
3. Given K_t, the firm maximizes its profits and input prices are given by w_t and rc_t and the rate of return on capital by r_t^k.
4. Given w_t, r_t, and c_t, g_{it} is the solution of the KF equation (14).
5. Given r_t, g_{it}, and B_t, the debt market (11) clears and $N_t = K_t - B_t$.

3.1 Equilibrium characterization

Several properties of the equilibrium are characterized with ease. We proceed first with the expert's problem. The use of log-utility implies that the expert consumes a constant share $\widehat{\rho}$ of its net wealth and chooses a leverage ratio proportional to the difference between the expected return on capital and the risk-free rate:

$$\widehat{C}_t = \widehat{\rho} N_t \tag{15}$$

$$\omega_t = \widehat{\omega}_t = \frac{1}{\sigma^2}\left(rc_t - \delta - r_t\right). \tag{16}$$

Second, rewriting the latter result, we get that the excess return on leverage,

$$rc_t - \delta - r_t = \sigma^2 \frac{K_t}{N_t},$$

depends positively on the variance of the aggregate shock, σ^2, and the leverage of the economy $\frac{K_t}{N_t}$. The higher the volatility or the leverage ratio in the economy, the higher the excess return the expert requires to isolate households from dZ_t. A positive capital growth rate shock, by increasing N_t relative to K_t, lowers the excess return.

Third, we can use the values of rc_t, L_t, and ω_t in equilibrium to get the wage $w_t = (1-\alpha) K_t^{\alpha}$, the rental rate of capital $rc_t = \alpha K_t^{\alpha-1}$, and the risk-free interest rate:

$$r_t = \alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t}. \tag{17}$$

This equation will play a key role in explaining our quantitative results. Since $K_t = N_t + B_t$, equations (15)-(17) depend only on the expert's net wealth N_t and debt B_t.

Fourth, we can describe the evolution of N_t:

$$\begin{aligned} dN_t &= \left[\left(r_t + \omega_t \left(rc_t - \delta - r_t\right)\right) N_t - \widehat{C}_t\right] dt + \sigma \omega_t N_t dZ_t \\ &= \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho} - \sigma^2 \left(1 - \frac{K_t}{N_t}\right) \frac{K_t}{N_t}\right) N_t dt + \sigma K_t dZ_t \end{aligned} \tag{18}$$

as a function only of N_t, B_t, and dZ_t. Equation (18) shows the nonlinear dependence of dN_t on the leverage level $\frac{K_t}{N_t}$. We will stress this point in the next pages repeatedly. For convenience, sometimes we will write

$$dN_t = \mu^N(B_t, N_t)dt + \sigma^N(B_t, N_t)dZ_t,$$

where $\mu^N(B_t, N_t) = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho} - \sigma^2 \left(1 - \frac{K_t}{N_t}\right) \frac{K_t}{N_t}\right) N_t$ is the drift of N_t and $\sigma^N(B_t, N_t) = \sigma K_t$ its volatility.

Fifth, we have from equation (12):

$$dB_t = \left(w_t + r_t B_t - C_t\right) dt = \left(\left(1 - \alpha\right) K_t^{\alpha} + \left(\alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t}\right) B_t - C_t\right) dt. \tag{19}$$

If we know C_t, N_t, and B_t, we can use (19) to find dB_t. With dB_t, we can calculate dK_t and all the other endogenous variables of the model follow directly (see Appendix A for a list of all the equilibrium conditions of the model to see this point). Hence, computing the equilibrium of this economy is equivalent to finding C_t and tracking the density $\{g_{it}(\cdot)\}_{t \geq 0}$ for $i \in \{1, 2\}$ that determines it.

3.2 The DSS of the model

Now, we describe the DSS of the model where there are no capital growth rate shocks, but we still have idiosyncratic household shocks. Thus, we set $\sigma = 0$ in the law of motion for the expert's net wealth (18) to find:

$$dN_t = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho}\right) N_t dt. \tag{20}$$

Since the drift of N_t, $\mu^N(B, N) = \left(\alpha K^{\alpha-1} - \delta - \widehat{\rho}\right) N$, must be zero in a DSS (recall that we drop the t subindex to denote the DSS value of a variable), we get $K = \left(\frac{\widehat{\rho}+\delta}{\alpha}\right)^{\frac{1}{\alpha-1}}$.

With this result, the DSS risk-free interest rate (17) equals the return on capital and the rental rate of capital less depreciation:

$$r = r_t^k = rc_t - \delta = \alpha K_t^{\alpha-1} - \delta = \widehat{\rho}. \tag{21}$$

As mentioned above, this condition forces us to have $\widehat{\rho} < \rho$. Otherwise, the households would

accumulate too many bonds and the DSS would not be well-defined.

Finally, the dispersion of the idiosyncratic shocks determines the DSS expert's net wealth:

$$N = K - B = K - \int a dG\left(a, z\right),$$

a quantity that, unfortunately, we cannot compute analytically.

3.3 The SSS of the model

An SSS is defined as a density $g^{SSS}(\cdot)$ and an equity N^{SSS} that remain invariant when the realization of the aggregate shock is zero. Let $\Gamma_\sigma(g(\cdot), N, W)$ be the law of motion of the economy given an aggregate capital volatility σ and a realization of the Brownian motion W. More precisely, $\Gamma_\sigma(\cdot,\cdot,\cdot)$ is an operator that maps income-wealth densities $g(\cdot)$ and equity levels N into changes in these variables:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \begin{bmatrix} g_{t+\Delta t}(\cdot) - g_t(\cdot) \\ N_{t+\Delta t} - N_t \end{bmatrix} = \Gamma_\sigma(g_t(\cdot), N_t, W_t).$$

The SSS, therefore, solves:

$$\Gamma_\sigma(g^{SSS}(\cdot), N^{SSS}, 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In general, we will have multiple SSSs that solve the previous functional equation. Indeed, several of them will appear in our quantitative exercise.

The difference between the SSS and the DSS is that the former is the steady state of an economy where individual agents make their decisions taking into account aggregate risks ($\sigma > 0$) but no shock arrives along the equilibrium path, whereas, in the latter, agents understand that they live in an economy without aggregate risks ($\sigma = 0$). The DSS is, then,

$$\Gamma_0(g^{DSS}(\cdot), N^{DSS}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

4 Solution

Our discussion of equation (19) highlighted the role of finding the households' aggregate consumption, C_t, to compute the equilibrium of the economy given some structural parameter values $\Psi = \{\alpha, \delta, \sigma, \widehat{\rho}, \rho, \gamma, z_1, z_2, \lambda_1, \lambda_2\}$.

To do so, we follow Krusell and Smith (1998) and assume that, when forming their expectations, households only use a finite set of n moments of the cross-sectional distribution of assets instead of the complete distribution. In contrast to Krusell and Smith (1998), in which the income-wealth distribution is the only endogenous state variable, here the expert's net wealth

N_t is also a state variable. At the same time, we do not have any exogenous state variable, as $K_t = N_t + B_t$ instantaneously incorporates the capital growth rate shocks.

For ease of exposition, we discuss the case with $n = 1$. At the cost of heavier notation, all the techniques can be trivially extended to the case with $n > 1$. More concretely, households consider a *perceived law of motion* (PLM) of aggregate debt:

$$dB_t = h\left(B_t, N_t\right) dt, \tag{22}$$

where $h\left(B, N\right)$ is the conditional expectation of dB_t given available information (B_t, N_t):

$$h\left(B_t, N_t\right) = \frac{\mathbb{E}\left[dB_t | B_t, N_t\right]}{dt},$$

instead of the exact law of motion (19). We borrow the term PLM from the learning literature (Evans and Honkapohja, 2001) to accentuate that we allow $h\left(\cdot, \cdot\right)$ to be a general function, and not just a polynomial function, perhaps with state-dependent coefficients, as in Krusell and Smith (1998). In Subsection 4.3, we propose a methodology in which the functional form $h\left(\cdot, \cdot\right)$ is obtained by applying machine learning to simulated data. This methodology will let the PLM approximate, arbitrarily well, equation (19). This extra flexibility is key given the complex nonlinearities present in laws of motion of N_t, equation (18), and B_t, equation (19).

Given the PLM, the household's problem has an associated Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{aligned} \rho V_i\left(a, B, N\right) &= \max_c \frac{c^{1-\gamma} - 1}{1-\gamma} + s\frac{\partial V_i}{\partial a} + \lambda_i\left[V_j(a, B, N) - V_i(a, B, N)\right] \\ &\quad + h\left(B, N\right)\frac{\partial V_i}{\partial B} + \mu^N\left(B, N\right)\frac{\partial V_i}{\partial N} + \frac{\left[\sigma^N\left(B, N\right)\right]^2}{2}\frac{\partial^2 V_i}{\partial N^2}, \end{aligned} \tag{23}$$

where $i \neq j = 1, 2$, and where we use the shorthand notation $s = s\left(a, z, N + B, G\right)$ from (8). Notice how the HJB incorporates $h\left(B, N\right)$. Equation (23) complements the equilibrium conditions (30)-(38) by making the problem of the household explicit.

4.1 An overview of the algorithm

Our algorithm to find $h(B, N)$ in (22) proceeds according to the following iteration:

1) Start with h_0, an initial guess for h.

2) Using the current guess for h, solve for household consumption, c, in the HJB equation (23). This solution can be obtained by using an upwind finite differences scheme described in Appendix A (although other numerical algorithms, such as a meshfree method or a deep neural network, can be used when the number of state variables is high).

3) Construct a time series for B_t by simulating the cross-sectional distribution over time. Given B_t, we can find N_t and K_t using equations (18) and (37).

4) Use a universal nonlinear approximator to obtain h_1, a new guess for h.

5) Iterate steps 2-4 until h_n is sufficiently close to h_{n-1} given some pre-specified norm and tolerance level.

Steps 1-5 show that our solution has two main differences with respect to the original Krusell-Smith algorithm: the use of continuous time and our employment of a universal nonlinear approximator to update the guess of the PLM. Both differences deserve some explanation.

4.2 Continuous time

Krusell and Smith (1998) wrote their model in discrete time. Our continuous-time formulation, while not changing any fundamental feature of the model, enjoys several advantages (Achdou et al., 2017; Nuño and Thomas, 2016). First, continuous time naturally generates sparsity in the matrices characterizing the transition probabilities of the discretized stochastic processes. Intuitively, continuously moving state variables such as wealth drift an infinitesimal amount in an infinitesimal unit of time. Thus, in an approximation that discretizes the state space, households reach only states that directly neighbor the current state. Second, the optimality characterizing consumption has a simpler structure than in discrete time:

$$c_i^{-\gamma} = \frac{\partial V_i}{\partial a}. \tag{24}$$

Third, it is easier to capture occasionally binding constraints such as equation (9) in continuous time than in discrete time as the optimality condition (24) for consumption holds with equality everywhere in the interior of the state space. Fourth, the dynamics of the cross-sectional wealth distribution are characterized by the KF equation (14). The discretization of this equation yields an efficient way to simulate a time series of the cross-sectional distribution (although this can also be performed in discrete time, as in Ríos-Rull 1997 and Reiter 2009, at some cost).

We simulate T periods of the economy with a constant time step Δt. We start from the initial income-wealth distribution at the DSS (although we could pick other values). A number of initial samples are discarded as a burn-in. If the time step is small enough, we have

$$B_{t_j+\Delta t} = B_{t_j} + \int_{t_j}^{t_j+\Delta_t} dB_s = B_{t_j} + \int_{t_j}^{t_j+\Delta t} h\left(B_s, N_s\right) ds \approx B_{t_j} + h\left(B_{t_j}, N_{t_j}\right) \Delta t.$$

Our simulation $\left(\mathbf{S}, \widehat{\mathbf{h}}\right)$ is composed of a vector of inputs $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_J\}$, where $\mathbf{s}_j = \left\{s_j^1, s_j^2\right\} = \left\{B_{t_j}, N_{t_j}\right\}$ are samples of aggregate debt and the expert's net wealth at J random times $t_j \in [0, T]$,

and a vector of outputs $\widehat{\mathbf{h}} = \left\{\widehat{h}_1, \widehat{h}_2 ..., \widehat{h}_J\right\}$, where

$$\widehat{h}_j \equiv \frac{B_{t_j+\Delta t} - B_{t_j}}{\Delta t}$$

are samples of the growth rate of B_t. The evaluation times t_j should be random and uniformly distributed over $[0, T]$ as, ideally, samples should be independent.

4.3 Neural networks: A universal nonlinear approximator

In the original Krusell-Smith algorithm, the law of motion linking the mean of capital tomorrow and the mean of capital today is log-linear, with the coefficients in that function depending on the aggregate shock. This approximation is highly accurate due to the near log-linearity of their models in the vicinity of the DSS. Indeed, in such a model, the DSS and SSS almost coincide. But, as shown in equations (18) and (19), this linearity of the law of motion of the endogenous variables with respect to other endogenous variables does not extend to our model.

This nonlinear structure causes two problems. First, we face the *approximation* problem: we need an algorithm that searches for an unknown nonlinear functional instead of a simple linear regression with aggregate-state-dependent coefficients. Second, we need to tackle the *extrapolation* problem. While the theoretical domain of B_t and N_t is unbounded, practical computation requires limiting it to a compact subset of $\mathbb{R}^2$ large enough to prevent boundary conditions from altering the solution in the subregion where the ergodic distribution accumulates. However, precisely because we deal with such a large area, the simulation in step 3 of the algorithm in Subsection 4.1 never visits an ample region of the state space. Thus, the approximation algorithm should provide not only an accurate nonlinear approximation in the visited region, but also a "reasonable" extrapolation to the rest of the state space. We will return to what "reasonable" means in this context momentarily.

To address these two problems, we employ a nonlinear approximation technique based on neural networks. Our approach displays four strengths. First, the *universal approximation theorem* (Hornik et al. 1989; Cybenko 1989) states that a neural network with at least one hidden layer can approximate any Borel measurable function mapping finite-dimensional spaces arbitrarily well. In particular, the theorem does not require that the approximated function be differentiable and can handle cases with kinks and occasionally binding constraints.[9]

Second, the neural network coefficients can be efficiently estimated using gradient descent methods and back-propagation. This allows for easier coding and shorter implementation time than other approaches.

[9]Lusin's theorem states that every measurable function is a continuous function almost everywhere. Thus, we can approximate jumps in a finite number of points, but not functions with extremely intricate shapes. Those complicated functions, however, are unlikely to be of much relevance in solving DSGE models.

Third, neural networks are more economical, for middle and high dimensions, than other approximators. Barron (1993) shows that, under some technical conditions, a one-layer neural network achieves integrated square errors of order $O(1/n)$, where n is the number of nodes. In comparison, for series approximations (polynomials, spline, and trigonometric expansions), the integrated square error is of order $O(1/(n^{2/d})$ where d is the dimensions of the function to be approximated. These results are extended by Bach (2017) to cover nondecreasing positively homogeneous activation functions such as the rectified linear unit and to derive approximation and estimation errors. In other words: the "curse of dimensionality" does not apply to neural networks that approximate functions of a very wide class. This advantage is not present in our baseline model, with $d = 2$, but will appear in any extension with additional state variables. Even going to $d = 3$ or $d = 4$ saturates alternatives such as Chebyshev polynomials.[10]

Fourth, neural networks extrapolate outstandingly. This is, in practice, key. Neural networks have well-behaved shapes outside their training areas. In contrast, Chebyshev polynomials (or other series) more often than not display explosive behaviors outside the fitted area that prevent the algorithm from converging. Figures D.1 and D.2 in Appendix D show this disappointing behavior of an approximation to the PLM in our model with Chebyshev polynomials. Within the area of high density of the ergodic distribution, Chebyshev polynomials approximate the law of motion for aggregate debt well (compare them with panel (c) in Figure 2, obtained with our neural network). Unfortunately, Chebyshev polynomials start oscillating as soon as we abandoned the well-traveled area of the simulation and the approximation becomes worthless.

We can now briefly describe our neural network approximator of the PLM. For introductory treatments of this material, see Bishop (2006) and Goodfellow et al. (2016). A neural network with a single hidden layer $h(\mathbf{s};\theta)$ is a linear combination of Q fixed nonlinear basis (i.e., activation) functions $\phi(\cdot)$:

$$h(\mathbf{s};\theta) = \theta_0^2 + \sum_{q=1}^{Q} \theta_q^2 \phi\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s^i\right), \tag{25}$$

where $\mathbf{s}$ is a two-dimensional input and θ a vector of weights:

$$\theta = \left(\theta_0^2, \theta_1^2, ..., \theta_Q^2, \theta_{0,1}^1, \theta_{1,1}^1, \theta_{2,1}^1, ..., \theta_{0,Q}^1, \theta_{1,Q}^1, \theta_{2,Q}^1\right).$$

Thus, the neural network provides a flexible parametric function h that determines the growth rate of aggregate debt $\widehat{h}_j = h(\mathbf{s}_j;\theta)$, $j = 1, .., J$.

We pick, as an activation function, a *softplus* function, $\phi(x) = \log(1 + e^x)$ for a given input x. The softplus function has a simple sigmoid derivative, which avoids some of the problems caused by the presence of a kink in rectified linear units, a popular choice in other fields, while keeping

[10] Similarly, approaches, such as a Smolyak interpolation, that alleviate the "curse of dimensionality" in some problems are difficult to apply here because the shapes of the ergodic distribution are hard to characterize ex-ante. Neural networks are more resilient to sparse initial information about the solution of the problem.

an efficient computation and gradient propagation.

The size of the hidden layer is determined by Q. This hypercoefficient can be set by regularization or, in simple problems, by trial-and-error. In our case, we set $Q = 16$ because the cost of a larger hidden layer is small. The neural network (25) can be generalized to include additional hidden layers. In that case, the network is called a *deep* neural network. However, for the problem of approximating a two-dimensional function, a single layer is enough.

The weights θ are selected to minimize the quadratic error function $\mathcal{E}\left(\theta; \mathbf{S}, \widehat{\mathbf{h}}\right)$ given a simulation $\left(\mathbf{S}, \widehat{\mathbf{h}}\right)$:

$$\theta^* = \arg\max_{\theta} \mathcal{E}\left(\theta; \mathbf{S}, \widehat{\mathbf{h}}\right) = \arg\max_{\theta} \frac{1}{2} \sum_{j=1}^{J} \left\| h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j \right\|^2.$$

A standard approach to performing this minimization is the *batch gradient descent* algorithm. Appendix B describes the training of the network and how we handle possible local minima.

Finally, our algorithm is massively parallel, either in CPUs, GPUs, or FPGAs (and, in the middle run, in the new generation of dedicated AI accelerators such as TPUs specially designed for this class of problems). This is a most convenient feature for scaling and estimation.

5 Estimation

Once we have solved the model given some structural parameter values Ψ, the next step is to take the model to the data by letting observations determine Ψ. We will proceed in two stages. First, we will discuss the simple case where the econometrician has access to output data and wants to build the likelihood associated with it. Second, we will show the results of our estimation with real data.

5.1 Building the likelihood function

Let us assume that the econometrician has access to $D + 1$ observations of output, $Y_0^D = \{Y_0, Y_\Delta, Y_{2\Delta}, ..., Y_D\}$ at fixed time intervals $[0, \Delta, 2\Delta, .., D\Delta]$. The derivations below would be similar for observables other than output. Since we have one aggregate shock in the model (to capital), we can only use one observable in our likelihood. Otherwise, we would suffer from stochastic singularity. If we wanted to have more observables, we would need to either enrich the model with more shocks or introduce measurement shocks in the observables. In those situations, we might need to resort to a sequential Monte Carlo approximation to the filtering problem described by the associated Kushner-Stratonovich equation of our dynamic system (see, in discrete time, Fernández-Villaverde and Rubio-Ramírez, 2007).

The likelihood function $\mathcal{L}_D\left(Y_0^D|\Psi\right)$ for our observations of output has the form:

$$\mathcal{L}_D\left(Y_0^D|\Psi\right) = \prod_{d=1}^{D} p_Y\left(Y_{d\Delta}|Y_{(d-1)\Delta};\Psi\right),$$

where $p_Y\left(Y_{d\Delta}|Y_{(d-1)\Delta};\Psi\right)$, the conditional density function of $Y_{d\Delta}$ given $Y_{(d-1)\Delta}$, is equal to:

$$p_Y\left(Y_{d\Delta}|Y_{(d-1)\Delta};\Psi\right) = \int f_{d\Delta}(Y_{d\Delta},B)dB$$

given a density function for output and debt, $f_{d\Delta}(Y_{d\Delta},B)$, implied by the solution of the model. Our task is, then, to compute the sequences of conditional densities $p_Y\left(Y_{d\Delta}|Y_{(d-1)\Delta};\Psi\right)$ at the fixed time intervals $[0,\Delta,2\Delta,..,D\Delta,]$.

To do so, we obtain the diffusion of $Y_t = (B_t+N_t)^\alpha$. Applying Itô's lemma, we get:

$$\begin{aligned} dY_t &= \frac{\partial\left(B+N\right)^\alpha}{\partial B}dB_t + \frac{\partial\left(B+N\right)^\alpha}{\partial N}dN_t + \frac{1}{2}\frac{\partial^2\left(B+N\right)^\alpha}{\partial N^2}\sigma^2\left(B+N\right)^2 dt \\ &= \mu^Y\left(B_t,Y_t\right)dt + \sigma_t^Y\left(Y_t\right)dZ_{t,} \end{aligned} \tag{26}$$

where:

$$\begin{aligned} \mu^Y\left(B_t,Y_t\right) &= \alpha Y_t^{\frac{\alpha-1}{\alpha}}\Bigg\{h(B_t,Y_t^{\frac{1}{\alpha}}-B_t) + \alpha Y_t + \left[\frac{\left(\alpha-1\right)\sigma^2}{2}-\delta\right]Y_t^{\frac{1}{\alpha}} \\ &\quad - \left(\alpha Y_t^{\frac{\alpha-1}{\alpha}} - \delta - \sigma^2\frac{Y_t^{\frac{1}{\alpha}}}{Y_t^{\frac{1}{\alpha}}-B_t}\right)B_t - \hat{\rho}\left(Y_t^{\frac{1}{\alpha}}-B_t\right)\Bigg\}, \end{aligned}$$

and $\sigma^Y\left(Y_t\right) = \alpha\sigma Y_t$.

With equation (26), the density $f_t^d(Y,B)$ follows the KF equation in the interval $[(d-1)\Delta, d\Delta]$:

$$\begin{aligned} \frac{\partial f_t}{\partial t} &= -\frac{\partial}{\partial Y}\left[\mu^Y(Y,B)f_t(Y,B)\right] - \frac{\partial}{\partial B}\left[h(B,Y^{\frac{1}{\alpha}}-B)f_t^d(Y,B)\right] \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial Y^2}\left[\left(\sigma^Y(Y)\right)^2 f_t(Y,B)\right]. \end{aligned} \tag{27}$$

At the beginning of the interval, we have $f_{(d-1)\Delta}(Y,B) = \delta\left(Y-Y_{(d-1)\Delta}\right)f_{(d-2)\Delta}(B|Y_{(d-1)\Delta})$, where $f_{(d-1)\Delta}(B|Y_{(d-1)\Delta})$ is the probability of B conditional on $Y = Y_{(d-1)\Delta}$:

$$f_{(d-2)\Delta}(B|Y_{(d-1)\Delta}) = \frac{f_{(d-2)\Delta}(Y_{(d-1)\Delta},B)}{f_{(d-2)\Delta}(Y_{(d-1)\Delta})} = \frac{f_{(d-2)\Delta}(Y_{(d-1)\Delta},B)}{\int f_{(d-2)\Delta}(Y_{(d-1)\Delta},B)dB},$$

if $d \geq 2$, $f_{-1}(B) = f(B)$ is the ergodic distribution of B, and $\delta(\cdot)$ is the Dirac delta function.

Lo (1988) pioneered the estimation of the likelihood of a continuous-time stochastic process

on discrete-time samples using the KF equation to characterize the transition density functions. That paper provides technical assumptions that have to be satisfied for the estimation. In our model, these conditions are met provided that $h(B, N)$ is twice continuously differentiable in B and N and three times continuously differentiable in Ψ, which is guaranteed if i) $h(B, N)$ is approximated using a neural network with our softplus activation function and ii) Ψ lies in the interior of a finite-dimensional closed and compact parameter space.[11]

A fundamental property of the operator in the KF equation (27) is that it is the adjoint of the infinitesimal generator employed in the HJB. The intuition for that result is that one can think about the dynamic choices of the agents implied by the HJB as a probability distribution of their future choices. Hence, agents' choices induce a distribution on observables, such as output, governed by the shocks of the model. There is an intimate link between optimal choices and likelihood functions.

This result is remarkable since it means that the solution of the KF equation amounts to transposing and inverting a sparse matrix that has already been computed when we solved the HJB. This provides a highly efficient way of evaluating the likelihood after the model is solved.[12]

In Appendix C, we briefly describe how to build the likelihood function of the model when we also add microeconomic observations from the cross-sectional distribution of assets.

5.2 Maximizing the likelihood

Once we have evaluated the likelihood, we can either maximize it or perform Bayesian inference relying on a posterior sampler. In this paper, for simplicity, we follow the former approach. Also, since we are dealing with a novel approach to the solution and estimation of models with heterogeneous agents, we keep the estimation as transparent as possible by fixing most of the structural parameters at conventional calibrated values for the U.S. economy. Also, we use only aggregate variables.

We will rely on U.S. quarterly output observations for 1984.Q1-2017.Q4, with bandpass filter keeping frequencies between 20 and 60 quarters (between 5 and 15 years). We start in 1984, as often done in the literature, to focus on capturing the dynamics that have governed aggregate fluctuations in the U.S. after the arrival of the Great Moderation (Galí and Gambetti, 2009). See also the updated evidence in Liu et al. (2018), who document how the Great Moderation has survived the 2007-2009 financial crisis. We bandpass the data to eliminate long-run trends and

[11] In our model, output is a continuous variable, whereas, in the data, is the cumulative production over a quarter. Thus, a more precise definition of the observable would be $Y^{agg}_{d\Delta} = \int_{(d-1)\Delta}^{d\Delta} Y_s ds$. In our paper, and given the low volatility of output for the range of considered parameters, this expression can be approximated with a high degree of accuracy by $Y^{agg}_{d\Delta} \approx Y_{d\Delta}\Delta$.

[12] If the KF would become numerically cumbersome in more general models, we could construct Hermite polynomial expansions of the (exact but unknown) likelihood as in Aït-Sahalia (2002). We could also consider methods of moments in continuous time such as those pioneered by Andersen and Lund (1997) and Chacko and Viceira (2003).

to skip the business cycle frequencies caused by productivity shocks and monetary policy shocks our model is not designed to account for. However, our methodology does not depend on this filtering and a richer model could be estimated with raw data without theoretical problems.

In terms of the fixed parameters, the capital share, α, is taken to be 0.35 and the depreciation rate of capital, δ, is 0.1 (all rates are annual). The discount rate, ρ, is set to 0.05. The risk aversion of the households γ is set to 2. These are standard values in the business cycle literature to match the investment-output ratio and the rate of return on capital.

The idiosyncratic income process is calibrated following our interpretation of state 1 as unemployment and state 2 as employment. The transition rates between unemployment and employment (λ_1, λ_2) are chosen such that (i) the unemployment rate $\lambda_2/(\lambda_1+\lambda_2)$ is 5% and (ii) the job finding rate is 0.3 at a monthly frequency or $\lambda_1 = 0.986$ at an annual frequency. These numbers describe the 'U.S.' labor market calibration in Blanchard and Galí (2010).[13] We normalize average income $\bar{z} = \frac{\lambda_2}{\lambda_1+\lambda_2} z_1 + \frac{\lambda_1}{\lambda_1+\lambda_2} z_2$ to 1. We also set z_1 equal to 71% of z_2, as in Hall and Milgrom (2008). Both targets allow us to solve for z_1 and z_2. We set the experts' discount rate $\widehat{\rho}$ to ensure that the leverage ratio K/N in the DSS is nearly 2, which is roughly the average leverage from a Compustat sample of nonfinancial corporations. Table 1 summarizes our calibration.

Parameter	Value	Description	Source/Target
α	0.35	capital share	standard
δ	0.1	capital depreciation	standard
γ	2	risk aversion	standard
ρ	0.05	households' discount rate	standard
λ_1	0.986	transition rate unemp.-to-employment	monthly job finding rate of 0.3
λ_2	0.052	transition rate employment-to-unemp.	unemployment rate 5%
z_1	0.72	income in unemployment state	Hall and Milgrom (2008)
z_2	1.015	income in employment state	$\mathbb{E}(y) = 1$
$\widehat{\rho}$	0.0497	experts' discount rate	$K/N = 2$

Table 1: Baseline parameterization.

We solve the model according to the algorithm in Section 4. We use four Monte Carlo simulations of 2.5 million years, each at a monthly frequency. We initialize the model at the DSS, and we disregard the first 500 years as a burn-in. We compute the PLM based on simulations on a region of the state space.[14]

[13] Analogously to Blanchard and Galí (2010, footnote 20), we compute the equivalent annual rate λ_1 as $\lambda_1 = \sum_{i=1}^{12} (1-\lambda_1^m)^{i-1} \lambda_1^m$, where λ_1^m is the monthly job finding rate.

[14] When forming expectations, households evaluate the PLM over the entire state space. Thus, the PLM is extrapolated over the regions of the state space not included in the support of the ergodic distribution. There is no guarantee that the dynamics of the model in the extrapolated region coincide with the ones expected in the PLM. Thus, as in Krusell and Smith (1998), our approximation can be interpreted as a self-justified equilibrium in which households' beliefs about the PLM coincide with the actual law of motion only in the equilibrium paths (Kubler and Scheidegger, 2018). Off-equilibrium, the PLM and the actual law of motion may diverge, but households never discover it, as this region is never visited.

Then, we evaluate the likelihood of the observations on U.S. output for different values of σ, the volatility of the aggregate shock and, therefore, the most interesting parameters in terms of the properties of the model. While doing so, we keep all the other parameter values fixed at their calibrated quantities. We maximize the likelihood function by searching on a grid between 0.013 and 0.015 with a step 0.0002 (we played extensively with σ values to determine the region of high likelihood before starting the grid search).

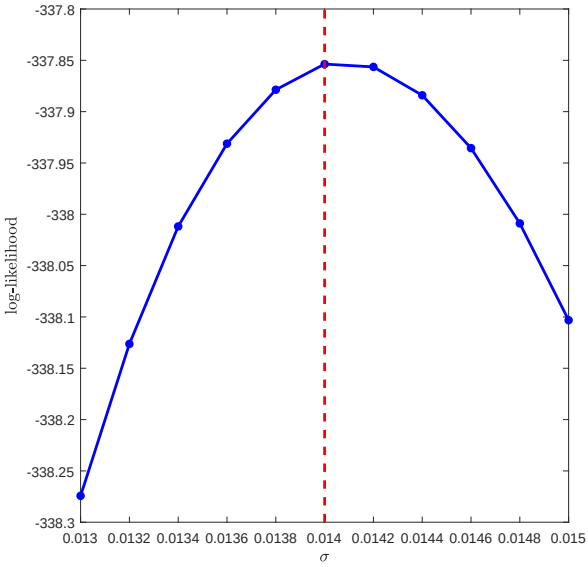

Figure 1: Log-likelihood for different values of σ and point estimate.

We plot the resulting log-likelihood in Figure 1. The point estimate, 0.0142, is drawn as a vertical discontinuous red line, with a standard error of 0.00011342, computed by the local derivative of the function. The smoothness of the plot confirms that our algorithm has successfully converged, since changes in one parameter value do not lead us into substantially different numerical solutions.

6 Quantitative results

This section presents the results generated by our solution algorithm with the parameter values from Section 5. We will report, first, the PLM for aggregate debt, $h(B, N)$ and assess its accuracy as a solution. Next, we will explore the phase diagram of the model and explain, through the dynamic responses of the model to an aggregate shock, why we find several SSS(s). After having discussed the convergence properties of the SSS(s), the random fluctuations around those, and documented the presence of time-varying aggregate risk in our economy, we will analyze the role of the value of σ in determining the multiplicity of SSS(s). We will close by looking at the aggregate ergodic distribution of debt and equity.

6.1 The PLM

Figure 2 reports the resulting PLM for aggregate debt, $h(B, N)$. Panel (a), at the top left, displays three transversal cuts of $h(B, N)$ along with a range of values of equity (N). The first cut fixes B at the high-leverage SSS (HL-SSS) value ($B^{HL} = 1.9641, N^{HL} = 1.7470$, with $K^{HL} = 3.7111$ and $\frac{K^{HL}}{N^{HL}} = 2.1243$), the second cut fixes B at an arbitrary high-leverage point ($B^* = 2.15, N^* = 1.5$, with $K^* = 3.65$ and $\frac{K^*}{N^*} = 2.4333$), and the third cut fixes B at the low-leverage SSS (LL-SSS; $B^{LL} = 1.0967, N^{LL} = 2.6010$, with $K^{LL} = 3.6977$ and $\frac{K^{LL}}{N^{LL}} = 1.4216$).

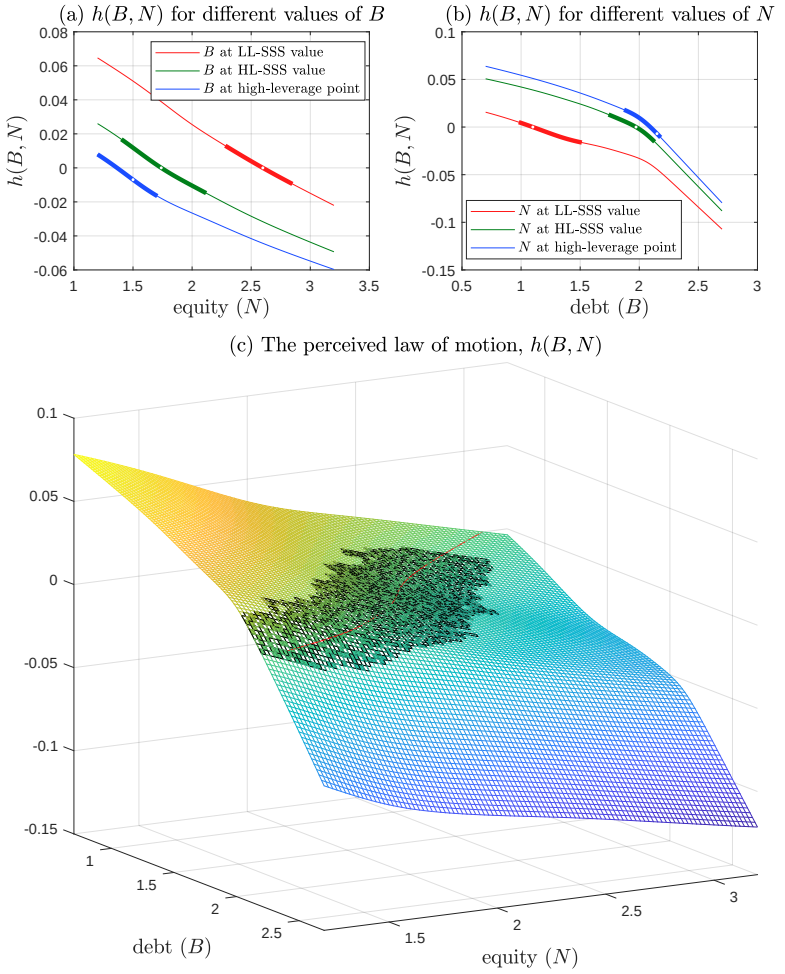

Figure 2: The PLM $h(B, N)$ and transversal cuts.

Note: White points in panels (a) and (b) indicate the LL-SSS, the HL-SSS, and an arbitrary high-leverage point. The thicker part of the lines in panels (a) and (b) and the shaded area in panel (c) display the region of the PLM visited in the ergodic distribution. The thin red line is the "zero" level intersected by the PLM.

The thicker part of the lines indicates the regions of the state space in which the ergodic distribution of aggregate variables is nonzero. The white point indicates, in the first top cut, the LL-SSS; in the second cut, the HL-SSS; and in the third cut, the high-leverage point described above. Panel (b), at the top right, follows the same pattern as panel (a), but switching the roles of equity (N) and debt (B). Finally, panel (c), at the bottom, shows the three-dimensional

representation of the PLM. The shaded area highlights the region of the PLM visited in the ergodic distribution with nontrivial positive probability. The thin red line is the "zero" level intersected by the PLM: to the right of the line, aggregate debt falls, and to the left, it grows.

Figure 2 demonstrates the nonlinearity of $h(B, N)$ even within the area of the ergodic distribution that has positive mass. The agents in our economy expect different growth rates of B_t in each region of the state space, with the function switching from concave to convex along the way. While this argument is clear from the shape of panel (c), it encodes rich dynamics. For instance, panel (b) shows how, as leverages increases, $h(B, N)$ becomes steeper and, in the ergodic distribution, more concave. Given the same level of debt, a higher level of leverage induces larger changes in the level of aggregate debt as the financial expert is exposed to more capital risk. This result will resurface several times in future paragraphs. Panel (c) also shows that, as intuition suggests, $h(B, N)$ is generally decreasing in debt and equity.

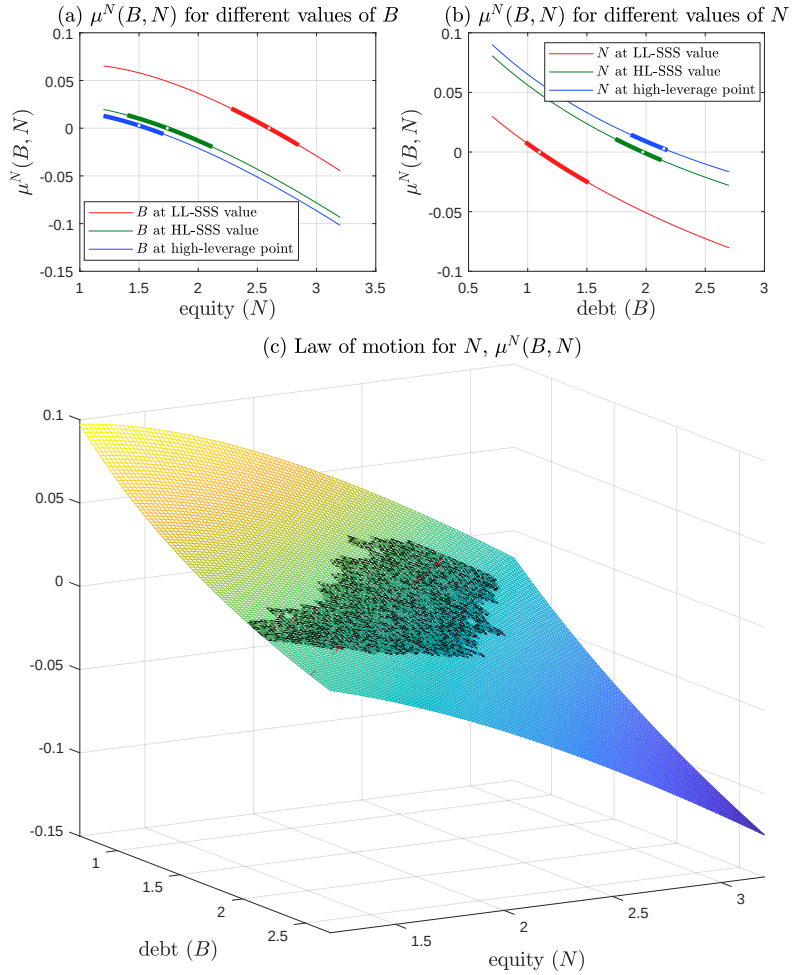

Figure 3: The law of motion $\mu^N(B, N)$ and transversal cuts.

Note: White points in panels (a) and (b) indicate the LL-SSS, the HL-SSS, and an arbitrary high-leverage point. The thicker part of the lines in panels (a) and (b) and the shaded area in panel (c) display the region of $dh(B, N)$ in the ergodic distribution. The thin red line is the "zero" level intersected by $\mu^N(B, N)$.

Figure 3 replicates Figure 2, except now for $\mu^N(B, N)$. Similar comments regarding the nonlinear structure of the solution apply here. For example, now, $\mu^N(B, N)$ becomes higher as a function of equity as the level of leverage falls.

The nonlinearity of the PLM confirms our conjecture that more traditional solution methods that rely on linear structures (conditional on aggregate shocks) might not be appropriate for solving this model. We further check this argument by looking at the forecasting capability of our PLM and comparing it with the forecasting capability of an alternative PLM computed using the Krusell and Smith (1998) algorithm (an OLS over a linear regression on endogenous state variables).

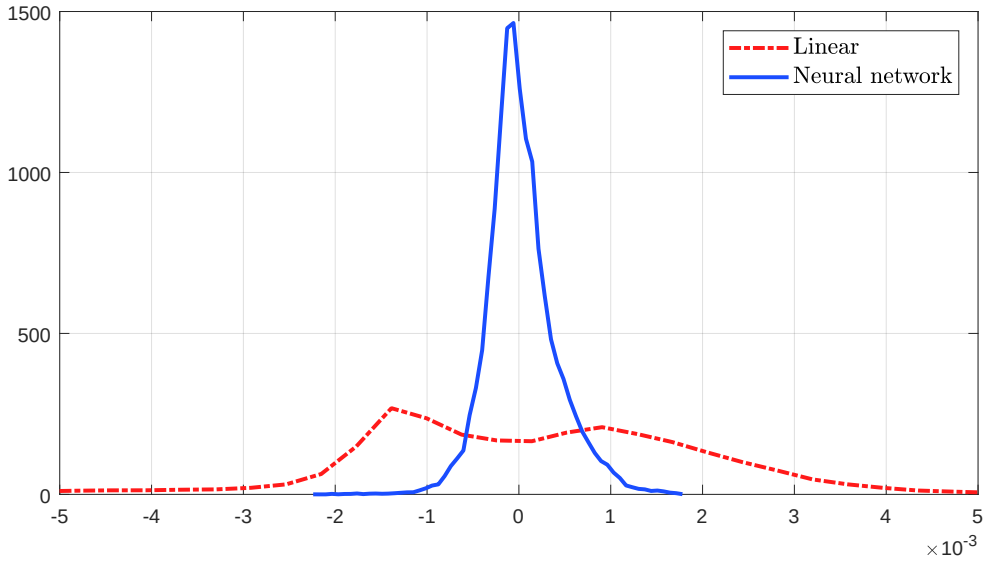

Figure 4: Forecasting error distribution at a one-month horizon, linear PLM (left) and neural network (right).

Figure 4 plots the histogram of forecasting errors at a one-month horizon (the time step in the simulation), with a continuous blue line representing the errors from our algorithm and the discontinuous red line the errors from a linear-in-endogenous state variables Krusell-Smith algorithm. The R^2 associated with the neural-network PLM is 0.9922, with an RMSE of 0.0004. The forecasting errors, furthermore, are nicely clustered around zero, with a mode roughly equal to zero. The Krusell-Smith algorithm produces an R^2 of 0.8275, considerably lower than typical values reported in the literature for standard heterogeneous agents models, and with an RMSE of 0.0021. Also, the forecasting errors are more volatile, skewed to the right, and without a mode at zero. We checked that adding more moments to the OLS regression does not help much in terms of accuracy.

In Appendix D, we discuss other alternative solution methods, such as Chebyshev polynomials, and argue that our method has advantages over them as well.

6.2 The phase diagram

Figure 5 plots the phase diagram of our model along with the aggregate debt (B) on the x-axis and equity (N) on the y-axis. The blue line in Figure 5 represents the loci of zero changes in debt, $h(B, N) = 0$. The line inherits the nonlinear dependence on B and N of the right-hand side of equation (19), the object $h(B, N)$ approximates. There is a convex segment for low levels of debt and a concave segment for high levels of debt. The discontinuous red line represents the loci of zero changes in the equity of the expert, $\mu^N(B, N)$. These two lines are the intersections of zero with $h(B, N)$ and $dN(B, N)$ in Figures 2 and 3. The arrows in Figure 5 indicate the movement of debt, B, and equity, N, when we are away from the blue and red lines.

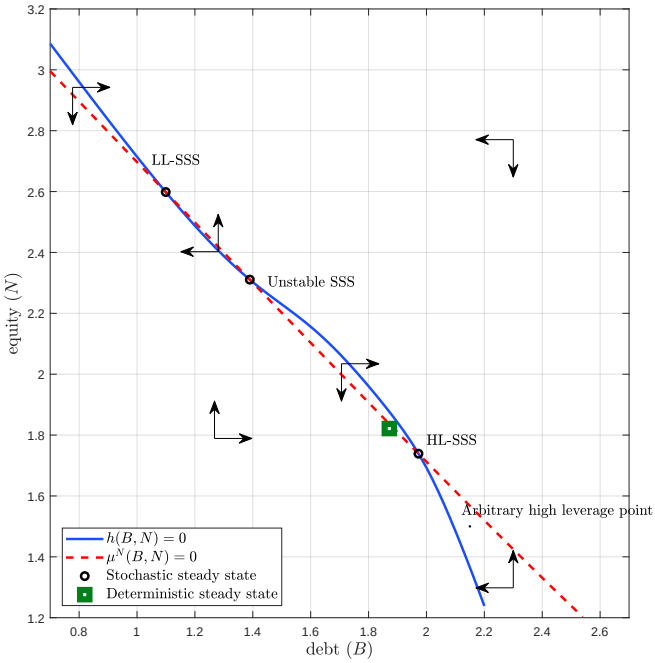

Figure 5: Phase diagram, DSS, and SSS(s).

The two lines intersect three times, defining three SSS(s). From the bottom right, the first intersection is the stable HL-SSS (recall, with $B^{HL} = 1.9641, N^{HL} = 1.7470$, and $K^{HL} = 3.7111$). This SSS is the closest to the DSS (green square). In comparison with the DSS ($B = 1.8718, N = 1.8215$, and $K = 3.6933$), the HL-SSS has 0.5% more capital, 4.9% more debt, and 4.1% less equity. The second intersection is at a middle-leverage SSS with less debt and more equity ($B^{ML} = 1.3897, N^{ML} = 2.3108$, and $K^{ML} = 3.7005$). This SSS is, however, unstable, and the dynamics of the economy quickly move away from it. Thus, we will not discuss it further. The final third intersection, at the top left, is the stable LL-SSS. Here, debt is much smaller ($B^{LL} = 1.0967$) and equity considerably higher ($N^{LL} = 2.6010$) than in the HL-SSS, yielding, however, a similar capital, $K^{LL} = 3.6977$. We also plot the point of high leverage that we use in Figures 2 and 3 to show, later, the behavior of the economy when leverage is high.

6.3 Differences in persistence at each stable SSS(s)

To explain why we have multiple SSS(s), we must first understand the differences in the persistence of the dynamic responses to a capital shock in each basin of attraction. To do so, Figure 6 displays the generalized impulse response functions (GIRFs) to a two-standard-deviations negative capital shock.[15]

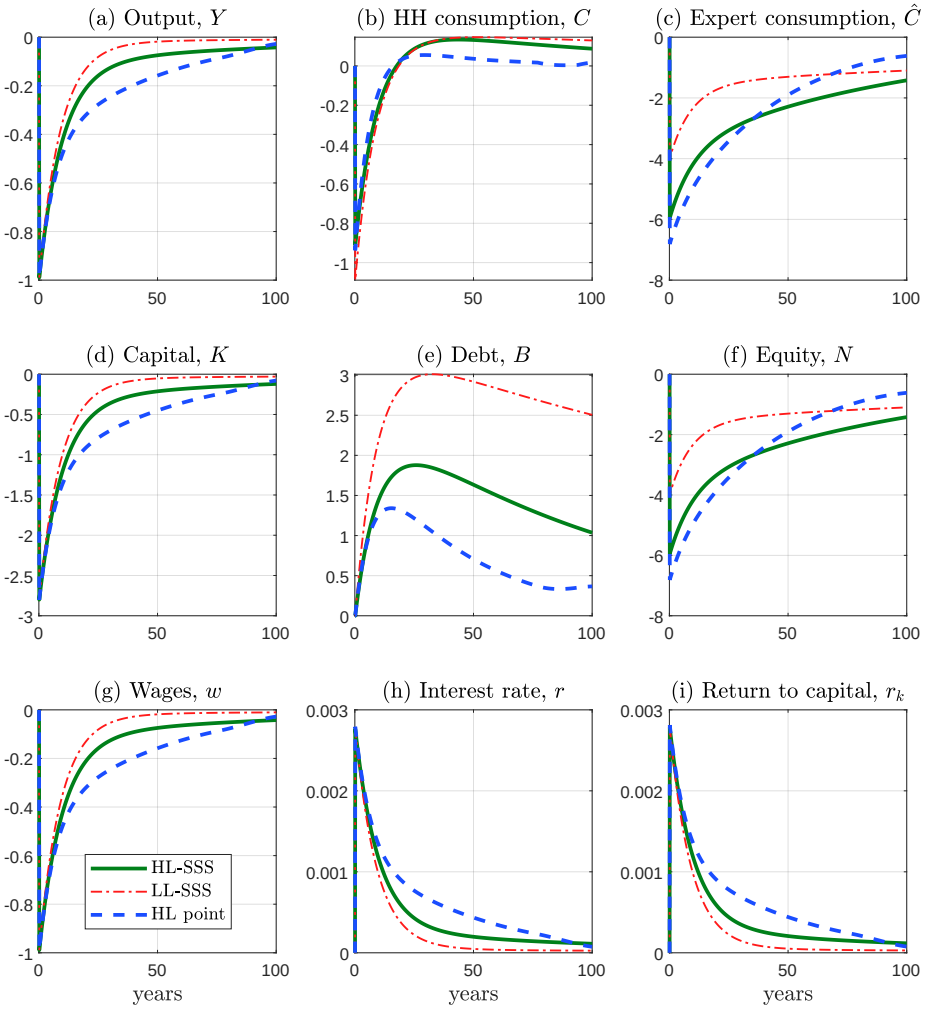

Figure 6: GIRFs for different initial states.

We plot three GIRFs: when the shock hits the economy at the HL-SSS (continuous green lines); when the shock hits the economy at the LL-SSS (discontinuous red lines); and when the shock hits the economy at the arbitrary high-leverage point defined in Figure 5 (discontinuous blue lines). Time units are years.

In all three cases, the shock destroys the same amount of capital (panel d) and output falls, at impact, the same (panel a). However, the higher the leverage, the larger the reduction of equity

[15]The word "generalized" is used because, in comparison with linear models, impulse responses in nonlinear models are i) state-dependent; ii) a nonlinear function of the size of the shock; iii) and nonsymmetric. Thus, we need to specify the size and sign of the shock and when this shock occurs.

at impact (panel f). To see this, we rewrite the law of motion for the expert's net wealth (18) as:

$$\frac{dN_t}{N_t} = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho}\right) dt + \sigma^2 B_t \frac{K_t}{N_t^2} dt + \sigma \frac{K_t}{N_t} dZ_t. \tag{28}$$

The volatility term in equation (28) depends positively on leverage, $\frac{K_t}{N_t}$: the expert must absorb the capital losses from a negative shock using a small net wealth. At the same time, equation (28) shows why the recovery rate of the expert's net wealth in panel (f) is roughly independent of leverage: the first drift term, $\left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho}\right)$, does not depend on leverage and the second drift term, $\sigma \frac{K_t}{N_t} dZ_t$, is negligible in size since it is multiplied by σ^2.

The expert also accumulates less debt after the negative shock when leverage is high (panel e). To analyze this point, we rewrite the law of motion for debt (19) as:

$$\frac{dB_t}{Bt} = (\alpha K_t^{\alpha-1} - \delta)dt - \sigma^2 \frac{K_t}{N_t} dt + (1-\alpha)\frac{K_t^{\alpha}}{B_t} dt - \frac{C_t}{B_t} dt. \tag{29}$$

The first drift term in equation (29), $\alpha K_t^{\alpha-1} - \delta$, does not depend on leverage. The second drift term, $\sigma^2 \frac{K_t}{N_t}$, is unimportant because of the presence of σ^2. The third term is smaller when leverage is high (and B_t is high for a given K_t). The fourth term, $\frac{C_t}{B_t}$, only depends weakly on leverage (we discuss why in the next paragraph). Putting the four terms together rationalizes why the growth rate of debt in panel (e) is lower when leverage is high.

Combining this latter result with the previous finding that the growth rate of equity is roughly independent of leverage, we get that capital (panel d) and output (panel a) recover more slowly when leverage is high. This gives us persistently lower wages (panel g) and a higher risk-free rate (panel b). Subsection 7.4 explains in detail how these changes affect the consumption of different households along the wealth distribution. Suffice it to say here that households experience a negative income effect (lower labor income), a positive income effect (higher bond income), and an intertemporal substitution effect (higher risk-free rate in their Euler equations). Given our parameter values, the net size of these opposite effects remains roughly unchanged as we vary the levels of leverage (panel b). For example, when leverage is high, the negative labor income effect is larger, but the positive risk-free rate income effect is bigger as well.

Figure 6 also explains why our economy diverges from a version of the model where the expert can issue outside equity without constraints. In such a case, the expert and the households share the capital risk. Thus, after a negative capital shock, we would not observe the large leverage movements in panel (f) that drive the nonlinear responses of other aggregate variables.[16]

Three final points about the GIRFs deserve discussion. First, the GIRFs are highly persistent. When leverage is high, even after 40 years, the economy is still around half a percentage point

[16]The importance of the financial friction on equity issuance would still go through qualitatively if the expert could issue some outside equity, but subject to some cap due to "skin-in-the-game" constraints. Quantitatively, though, the effect of capital shocks could be smaller.

below its pre-shock level. The dynamics of equity and debt accumulation propagate aggregate shocks in ways that are not present in models without financial frictions (see Subsection 6.7 for more details). Second, the state-dependence of the GIRFs induces time-variation in the volatility of output and other aggregate variables, a basic feature of U.S. data (Fernández-Villaverde and Guerrón-Quintana, 2020). Third, the two-standard-deviations shock is not large enough to send the economy away from the basins of attraction of each SSS. An even larger shock or a shock closer to the frontier between the two basins will induce a switch of basin, the endogenous regime-switching behavior we discussed in the introduction. These larger shocks induce a sharp nonlinearity, as the GIRFs will converge to the SSS of the new basin, not their origin. We will return to this point in Subsection 6.5.

6.4 Why do we have two stable SSS(s)?

We are now ready to discuss why we have two stable SSS(s). The key is the evolution of wages (panel g) in Figure 6. Because of the arguments in the previous subsection, when leverage is high, wages fall for a much longer period than when leverage is low. In the HL-SSS, it takes 19.08 years for 75% of the effect on wages of a negative aggregate shock to dissipate (and 30 years at the high-leverage point). In comparison, it takes 13.33 years in the LL-SSS. Similarly, the standard deviation of wages when the economy fluctuates around the basin of attraction of the HL-SSS is 0.0125, but 0.0108 around the basin of attraction of the LL-SSS.

The forecast of likely persistent falls in wages embedded in the PLM induces a stronger precautionary behavior by households. Since they want to smooth their consumption, households' demand for debt shifts to the right when leverage is high. At the same time, the supply of debt given by equation (16) is:

$$B_t = \left(1 + \frac{1}{\sigma^2}(rc_t - \delta - r_t)\right) N_t.$$

Thus, the debt market will clear with a reduction in r_t. At the HL-SSS, this reduction in r_t will also need to compensate for the small N_t (the HL-SSS is characterized by having low equity) and slightly lower rc_t (as total capital is a bit higher than at the LL-SSS).

Hence, we have shown why we have a fixed point at the HL-SSS. High leverage makes wages persistently lower after a negative capital shock. Persistently lower wages create a precautionary motive that lowers the risk-free rate and induces a high leverage by the expert. We only need to reverse the argument to show why we also have a fixed point at the LL-SSS. Low leverage makes wages recover quickly after a negative capital shock. The lower associated precautionary behavior means the risk-free rate is relatively high and that sustains the low leverage of the expert.

The high sensitivity of leverage to excess returns induced by $\frac{1}{\sigma^2}$ in equation (16) accounts for the large differences between the two SSS(s) in terms of how a roughly equal total wealth ($K^{HL} = 3.7111$ vs. $K^{LL} = 3.6977$) is allocated between experts ($N^{HL} = 1.7470$ vs. $N^{LL} = 2.6010$) and

households ($B^{HL} = 1.9641$ vs. $B^{LL} = 1.0967$). The excess return for the HL-SSS is 4.1636 basis points and, for the LL-SSS, 2.7864. This high sensitivity is the dual of the negligibility of the terms multiplied by σ^2 in equations (28) and (29).

6.5 Convergence to the SSS(s) and random fluctuations around them

How do we know that the two SSS(s) described above are stable? The state space $(g(\cdot), N)$ is infinite-dimensional and, hence, we cannot check convergence numerically for all possible initial states. Instead, we analyze convergence for densities visited in the ergodic distribution.

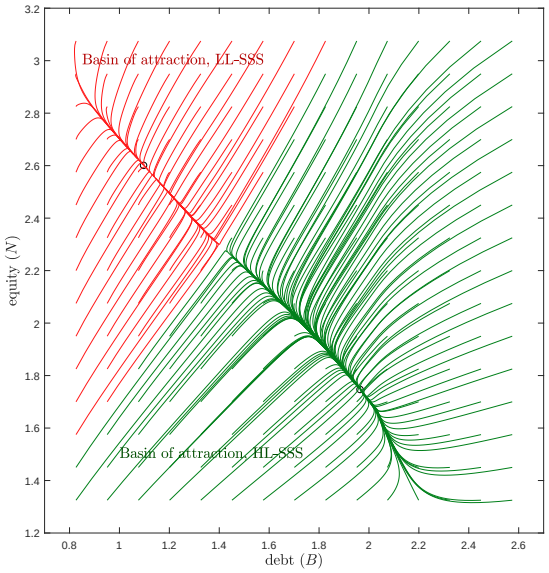

Figure 7: Convergence paths.

Figure 7 considers an array of different initial income-wealth densities and equity levels $(g_0(\cdot), N_0)$, selected from the simulations used to compute the aggregate ergodic distribution and analyze the transitional dynamics in the absence of aggregate shocks (agents continue forming their expectations assuming $\sigma > 0$). We plot, in red, the paths converging toward the LL-SSS and, in green, the paths converging toward the HL-SSS. In all cases, the economy converges to the SSS(s) –denoted by small circles– on the basin of attraction of the initial condition. However, the convergence path is plodding and it may take centuries.

Notwithstanding, we cannot rule out that, for other initial conditions, the model would not converge to an SSS. This limitation is related to the self-justified nature of the solution. The PLM is computed based on the income-wealth distributions visited along the ergodic distribution. One could potentially find a distribution that would lead to alternative dynamics.[17]

[17]Similarly, one could also find other equilibria beyond the one we compute (although, despite our efforts, we failed to do so). Recall that the multiplicity of SSS(s) is different from a possible multiplicity of equilibria: in our model, we are in one basin of attraction or another depending on the sequence of shocks the economy has experienced, but the equilibrium we compute is unique given the initial condition and sequence of shocks.

An interesting feature of the transitional dynamics is that equity and debt often overshoot their SSS values. For example, when the economy starts with low levels of equity and high debt, the financial expert issues even more debt for a while. Only as the expert accumulates wealth through excess returns undisturbed by shocks (we are in the deterministic convergence path) does equity grow and debt fall.

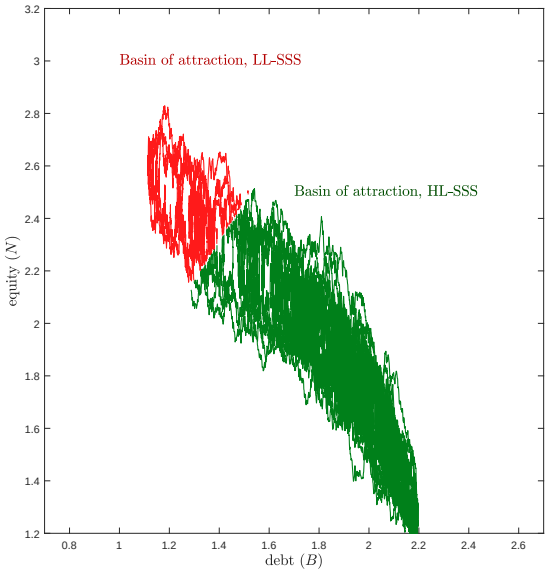

Figure 8: Simulation of equilibrium paths.

Figure 8 documents how the economy evolves around each SSS by plotting several random equilibrium sequences corresponding to different sequences of aggregate shocks. In red, we plot the paths when they are in the basin of attraction of the LL-SSS. In green, we plot the paths when they are in the basin of attraction of the HL-SSS.

Figure 8 shows how, along the equilibrium sequence of the economy, our economy has recurrent endogenous regime-switches, with their associated variations in the persistence, volatility, and skewness of aggregate variables. A path in the basin of attraction of the LL-SSS can be pushed toward the basin of attraction of the HL-SSS by a series of shocks that reduce equity and increase debt. Conversely, a series of shocks that increase equity and reduce debt can push a path from the basin of attraction of the HL-SSS toward the basin of attraction of the LL-SSS.

6.6 Endogenous aggregate risk

Table 2 reports the moments of the economy conditional on the basin of attraction at which the variables fluctuate (note that the ergodic distribution of the economy over time is the combination of the distributions at both basins; see Subsection 6.9 below). Figure 9 plots the distributions behind the moments in Table 2.

There are three main takeaways from Table 2. First, the mean of output is higher at the

	Mean	Standard deviation	Skewness	Kurtosis
$Y^{\text{basin } HL}$	1.5802	0.0193	0.0014	2.869
$Y^{\text{basin } LL}$	1.5829	0.0169	0.1186	3.0302
$r^{\text{basin } HL}$	4.92	0.3364	0.0890	2.866
$r^{\text{basin } LL}$	4.89	0.2947	-0.0282	3.0056
$w^{\text{basin } HL}$	1.0271	0.0125	0.0014	2.8691
$w^{\text{basin } LL}$	1.0289	0.0111	0.1186	3.0302

Table 2: Moments conditional on basin of attraction.

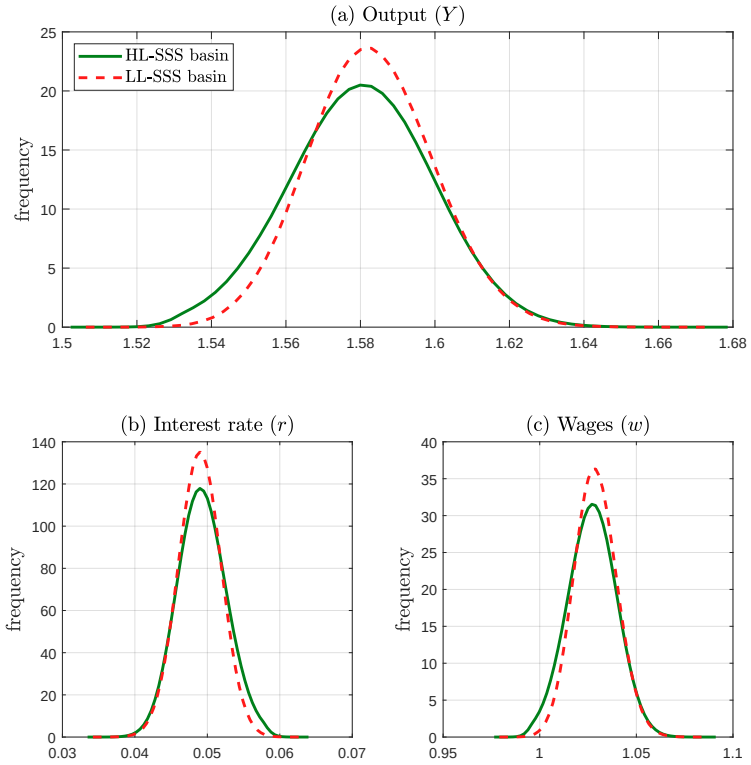

Figure 9: Ergodic distributions conditional on basin of attraction.

LL-SSS basin than at the HL-SSS. This is despite output being higher at the HL-SSS itself than at the LL-SSS ($Y^{HL} = 1.5824 > Y^{LL} = 1.5804$). The reversal is due to the negative skewness of the distribution of output at the basin of the HL-SSS in comparison with the positive skewness of the distribution of output at the basin of the LL-SSS (see panel a in Figure 9; Fernández-Villaverde and Guerrón-Quintana, 2020 discuss skewness in the main macro variables of the U.S. economy). The levels of very low output at the HL-SSS basin occur in the aftermath of large negative shocks to capital, as the economy struggles to recover as described in Subsection 6.4.[18] Second, the volatility (i.e., standard deviation) of output is higher in the basin of the HL-SSS:

[18]See Dew-Becker et al. (2019) for evidence that aggregate activity is negatively skewed. Since our model suggests that over the last several decades we have been at the HL-SSS basin, this is what we would expect to find in the data.

when leverage is high, the economy fluctuates more. The mechanism, outlined a few paragraphs back, is the higher persistence of capital in the basin of the HL-SSS after a shock. Third, there is a mild excess kurtosis of output at the HL-SSS.

Similar points can be made about the risk-free rate and the wage. The mean of the risk-free rate is slightly higher in the basin of the HL-SSS (4.92) than in the basin of the LL-SSS (4.88) even if the risk-free rate at the HL-SSS itself is lower than at the LL-SSS. This is because the basin of the HL-SSS includes an area of high risk-free rates (panel b in Figure 9) that appear after large negative shocks to capital. This phenomenon also appears as a higher standard deviation of r at the basin of the HL-SSS. Conversely, the mean of the wage is lower at the basin of the HL-SSS (despite being higher at the HL-SSS itself and the neighborhood around it) and more volatile because of the persistently low wages after large negative shocks (panel c in Figure 9).

6.7 Supercycles of borrowing and deleveraging

Figure 10 plots the histogram of the duration of spells of the economy around each SSS in a log scale.[19] The interaction of financial frictions and precautionary behavior leads to persistent middle- and long-run dynamics that are not present in conventional DSGE models, with supercycles of borrowing and deleveraging that may last for decades (and, on occasion, for centuries).

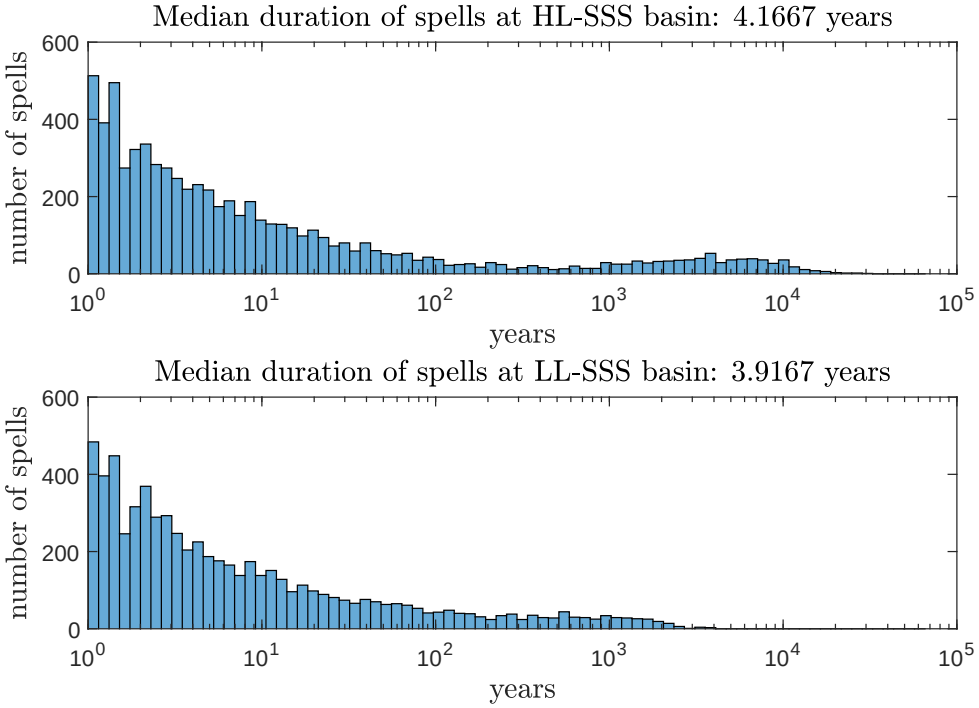

Figure 10: Spell durations at each SSS.

[19]To make this section more informative, we eliminate from our statistics and Figure 10 the spells with a duration of less than one year. When the value of the state variables is close to the frontier between the two basins of attraction, there are many extremely short spells as the economy jumps from one basin to the other after a shock before the convergence dynamics of the model push the state variables closer to the SSS.

The median duration of a spell around the HL-SSS is 4.167 years and, around the LL-SSS, 3.917 years. However, the distribution of spells at the HL-SSS has a much thicker right tail, with many spells lasting centuries. Thus, the economy spends 86.738% of its time (86.706% if we condition on spells lasting more than one year) in the basin of attraction of the HL-SSS.

Why does the economy stay longer, on average, around the HL-SSS? For our parameter values, i) the state variable values where the financial expert wants to lever aggressively cover a large region; and ii) the level of exogenous aggregate and idiosyncratic risk is sufficiently high as to induce households to demand enough debt.

Figures 6 and 10 also show how our model generates supercycles of borrowing and deleveraging, like those documented by Reinhart and Rogoff (2009). Imagine, for instance, an economy that is pushed to the LL-SSS basin because of a significant deleveraging (this can be due to widespread defaults, a massive redistribution generated by a war or a considerable policy shock, or high inflation). After a series of negative capital shocks that send the economy back to the HL-SSS basin, the economy will experience a protracted process of progressively higher wealth inequality, increasing debt and leverage, falling risk-free rates, and larger responses to capital shocks, interestingly, even if the shocks that pushed the economy to this basin disappear.

This process resembles the experience of many rich economies after 1945. For example, the U.S. came out of World War II with an extremely low level of household and corporate debt, the fruit of pervasive defaults and the lack of new borrowing during the Great Depression and the war (McKinsey Global Institute, 2010, pp. 71-73).[20] This situation lasted for a couple of decades, until leverage started rising quickly during the 1970s. Our model attributes such an increase to an erosion in the expert's net wealth relative to total capital. This erosion was caused, in the data, by a combination of high workers' bargaining power, high capital income taxes, and the oil shocks of the early 1970s (see Blanchard, 1997, and Drautzburg et al., 2017, for empirical details). Once the economy was in the new basin, the leverage process reinforced itself (even after the shocks mentioned above dissipated), and higher leverage was accompanied by higher wealth inequality, falling risk-free rates, and heightened fragility of the economy to adverse shocks. A similar picture holds for the U.K. (McKinsey Global Institute, 2010, pp. 74-75), although in this case, inflation played a crucial role in the initial reduction of the real value of debt.

6.8 The role of exogenous aggregate volatility

Our previous discussion highlighted how the excess return sends the economy toward a high- or a low-leverage region. Figure 11 pushes this argument further by exploring the role of σ, the exogenous aggregate volatility, in determining the different SSS(s). Each panel plots, for a different value of σ, the phase diagram of the economy following the same convention as in Figure

[20] Our model focuses on private debt. But even if we consider government debt, the U.S. had low levels of aggregate debt at the end of 1945.

5. Here, as in all the other exercises of the paper where we perform comparative statics on the values of one or two parameters, we keep all the other parameter values constant.

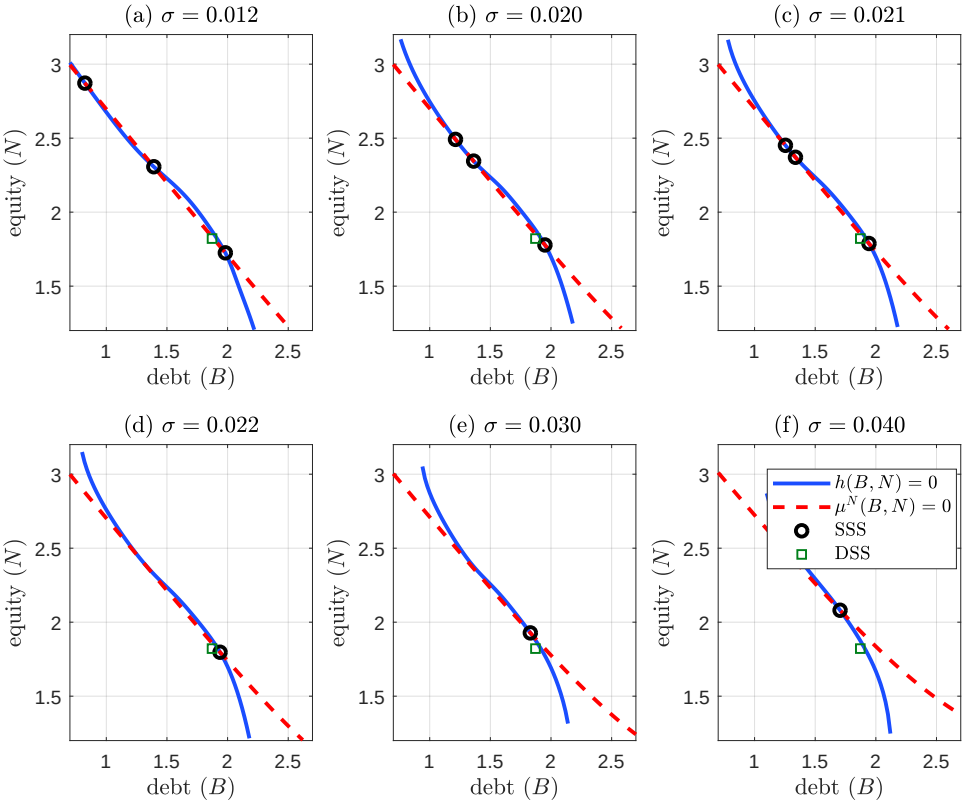

Figure 11: Phase diagram as a function of σ.

For low values of σ, the precautionary motive of households is mild and, therefore, we can sustain the LL-SSS (in addition to the HL-SSS). However, as we increase σ, the precautionary motive becomes stronger and $h(B, N) = 0$ is curvier, until we get to a bifurcation (i.e., a point where the system has a sudden topological change in its behavior) and the LL-SSS disappears. Even with high risk-free rates, households demand few bonds. Similarly, as σ grows, the HL-SSS moves to the left (i.e., less debt and more equity) as the expert is exposed to additional capital risk. This effect becomes sufficiently strong that the HL-SSS, instead of being to the right of the DSS (i.e., more debt and more equity than the DSS because of the higher excess return induced by precautionary savings), crosses to the left of the DSS.

Figure E.3 in Appendix E plots the values of the LL-SSS, the unstable SSS, and the HL-SSS (plus, for reference, the DSS) as a function of σ. Over there we can see how the leverage in the HL-SSS is a negative function of σ, a roughly constant function in the unstable SSS, and an increasing function in the LL-SSS (until the additional SSS(s) disappear). The mechanism for these three slopes is the same as the one discussed above. In the HL-SSS, as σ grows, the expert wants to unload some of the capital risk by reducing its leverage. In comparison, in the LL-SSS, households demand more debt as σ increases.

6.9 The aggregate ergodic distribution of debt and equity

Panel (a) of Figure 12 displays the aggregate ergodic joint distribution of debt and equity $F(B,N)$. This distribution is defined, for any subset Ω of the state space, as:

$$\mathbb{P}\{(B,N)\in\Omega\}=\int_{\Omega}dF.$$

The ergodic distribution is not obtained directly from the PLM, but from the simulation of the paths of the income-wealth distribution and equity. The PLM is employed, instead, in the HJB equation (23) to obtain the optimal consumption decision rules of individual households.

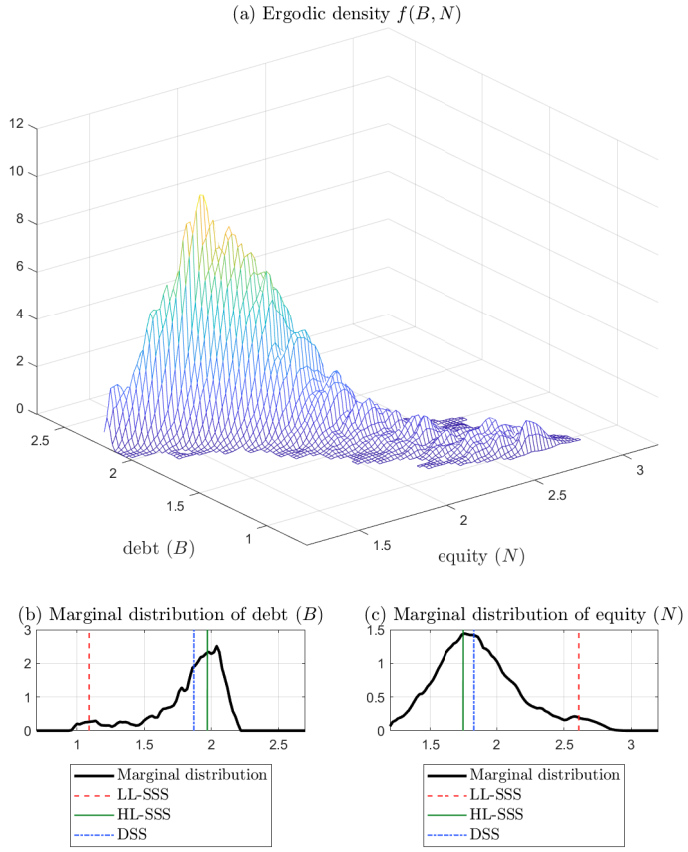

Figure 12: Ergodic distributions. Lighter colors indicate higher probability.

Panels (b) and (c) of Figure 12 plot the marginal ergodic distributions for debt and equity. These two panels show how the economy spends most of the time in a region with debt levels between 1.5 and 2.2 and equity between 1 and 2.2. Those values correspond to the neighborhood of the HL-SSS. In comparison, the neighborhood of the LL-SSS appears much less often. As seen in Figure 10, the desire of households to accumulate more debt lowers the height of this second peak as there is a strong force for reversion toward higher levels of leverage. We will see momentarily how this result changes as we vary the degree of heterogeneity among households.

Note, as well, how the marginal distribution of debt is much more concentrated than the

marginal distribution of equity. The ergodic distribution has a substantial tail in areas of high equity and low debt, creating an intriguing asymmetry in aggregate dynamics.

7 The role of household heterogeneity

This section explores how heterogeneity among households drives our results. We start by comparing, in the top row of Figure 13, the wealth distribution in the DSS (discontinuous blue line), the HL-SSS (continuous green line), and the LL-SSS (discontinuous red line). In the left panel, we plot the distributions for low-z households (with circles denoting the mass at zero assets). In the right panel, we plot the distributions for high-z households.

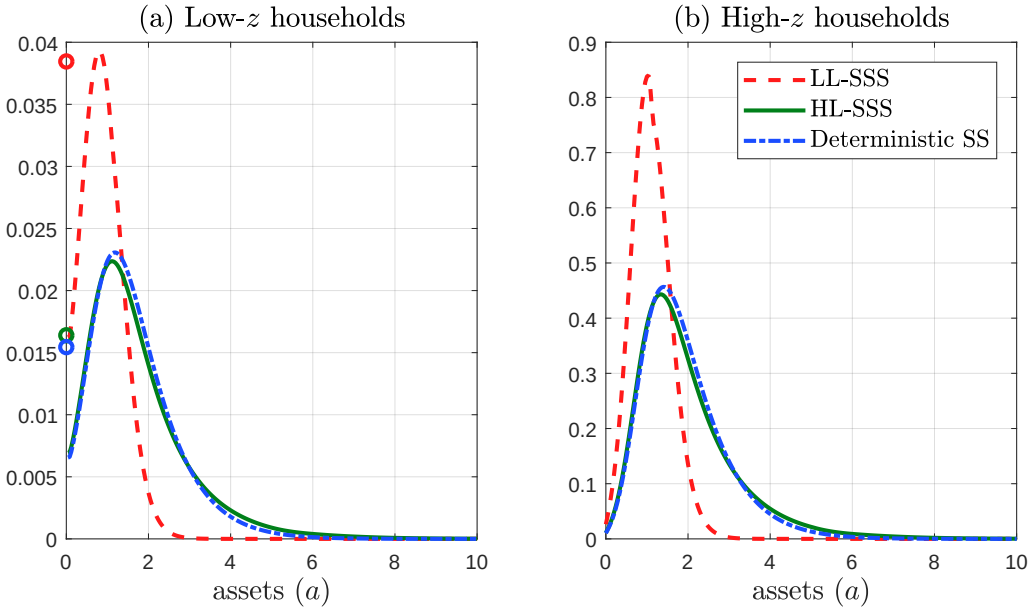

Figure 13: Wealth distribution in the DSS and SSS.

The wealth distribution slightly shifts to the left as we move from the DSS to the HL-SSS, but shifts rather dramatically, also to the left, as we travel from the HL-SSS to the LL-SSS. The reason behind these moves is the changes in endogenous aggregate risk we have discussed before. The higher volatility and persistence of wages in the HL-SSS generate a thicker right tail than at the DSS. These movements lead to substantial differences in the Gini coefficient of wealth: 0.28977 in the DSS, 0.24398 in the LL-SSS, and 0.31968 in the HL-SSS.

Having more wealth also means that the average welfare of the households at the HL-SSS is higher. Households are willing to give up, on average, 0.033633% of their consumption to (instantaneously) move from the DSS to the HL-SSS. However, we need to compensate them, also on average, 0.33983% of their consumption to go to the LL-SSS. These numbers must be read carefully, though. The value functions of the households at the LL-SSS are above the value functions of those at the HL-SSS for all asset values because the economy is less volatile and

the utility function is concave (see the plot of the value functions in Figure F.7 in Appendix F). However, we have more households with larger assets at the HL-SSS.

Next, we show how the results of Section 6 change as we modify the forces driving the wealth distributions in Figure 13. We will answer this question in four steps. First, we will vary the level of idiosyncratic risk among households. Second, we will change aggregate and idiosyncratic risk simultaneously. Third, we will show how the GIRFs of the economy depend on the degree of heterogeneity among households. Fourth, we will explore how the differences in consumption decisions and the variation in distributions over time account for the differences in the GIRFs. We will conclude with a brief comment on our approximation of the PLM.

7.1 Varying idiosyncratic labor risk

Figure 14 plots the phase diagram of the model for nine different values of the unemployment income, z_1 (still keeping the ergodic mean of z equal to 1) from 0.67 to 0.92. Each panel follows the same convention as in Figure 5. In particular, we only plot the segments of $h(B, N)$ and $\mu^N(B, N)$ visited in the ergodic distribution. Figure 14 is different from Figure 11 because now we keep σ constant as we change idiosyncratic risk.

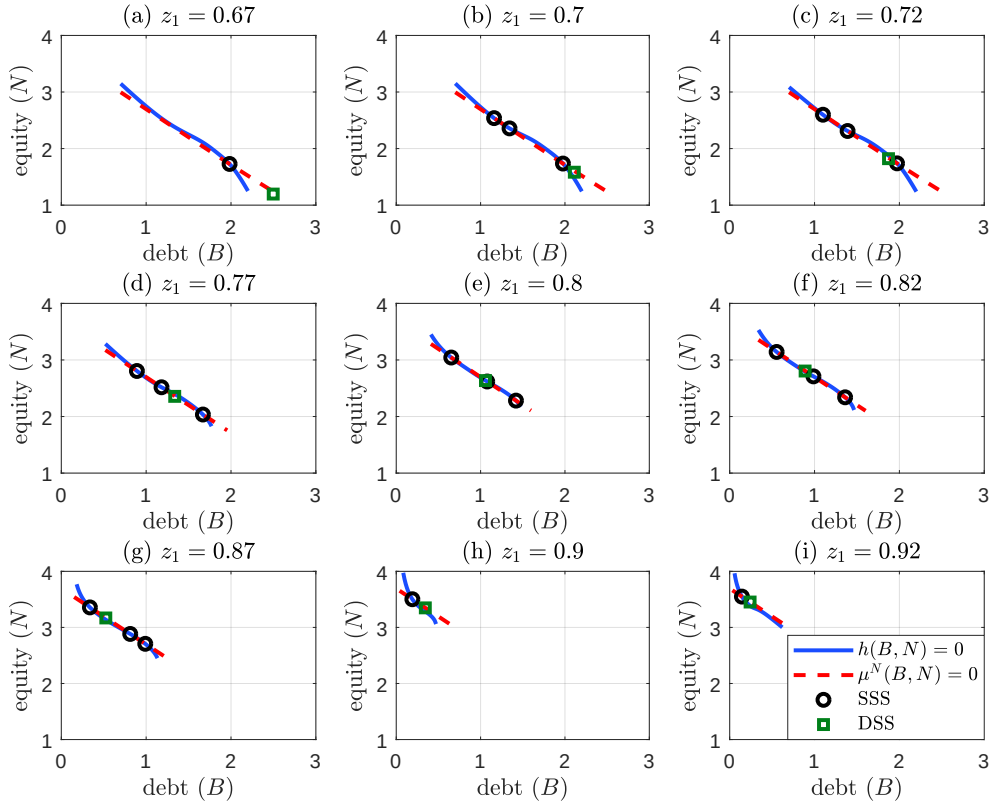

Figure 14: Phase diagram as a function of z_1.

For low levels of z_1, the only SSS is the HL-SSS (see the left and center panels at the top

row). Idiosyncratic risk is so high that households demand enough debt to sustain only one SSS, which, besides, has more debt than the DSS. As z_1 increases, we cross a threshold around 0.69 and we find three SSS(s), with the same interpretation as in Figure 14. As z_1 continues to rise, the HL-SSS moves to the left of the DSS since households demand less debt to self-insure against idiosyncratic risk. See the three panels of the center row. By the time z_1 reaches 0.9, the HL-SSS has disappeared: the precautionary demand for debt by households is now so weak that only the LL-SSS survives.

Figure 15 draws the ergodic distributions of equity and debt as z_1 varies. For low levels of z_1, most of the ergodic mass accumulates in the region of high debt and low equity. As z_1 increases, the ergodic mass spreads toward the upper left corner, first slowly, but gathering steam by the time we reach $z_1 = 0.85$. At this level, there is a bifurcation and the region around the LL-SSS becomes predominant and the higher leverage region eventually disappears. This change in the ergodic distribution is crucial for aggregate fluctuations since, as we saw in Figure 6, the responses of the economy to a capital shock heavily depend on leverage.

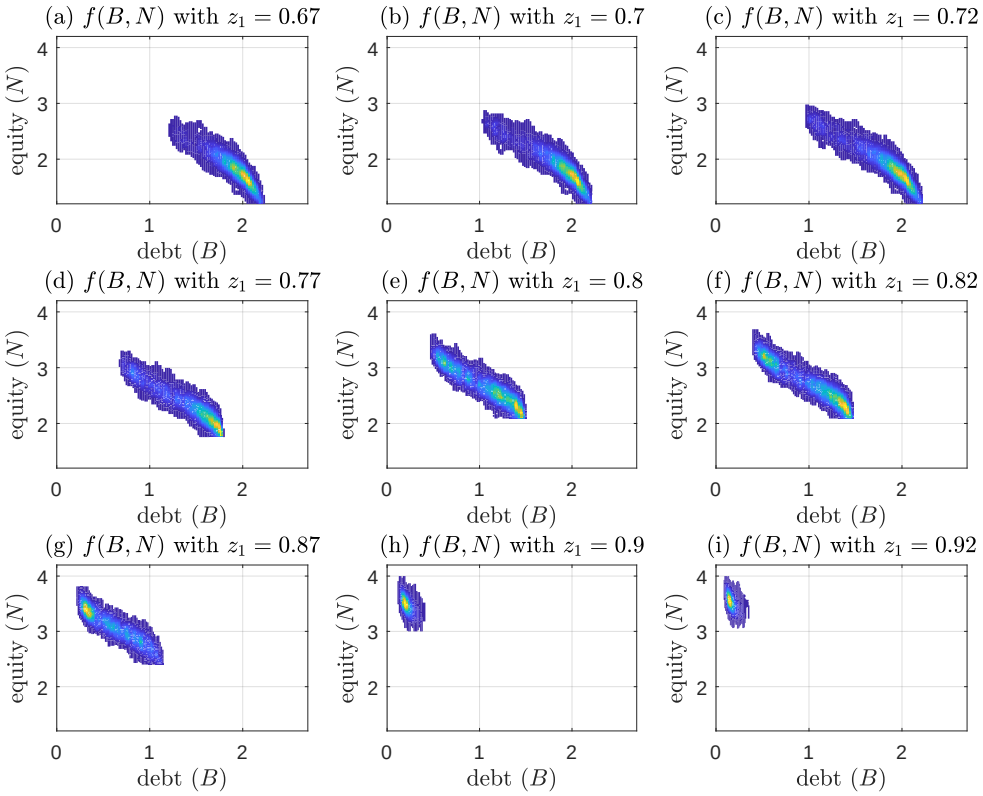

Figure 15: $f(B, N)$ as a function of z_1. Lighter colors indicate higher probability.

A different view of the same results appears in Figure 16, where we plot the marginal densities of debt, equity, and capital as we change z_1. Here, we appreciate the very low debt for high z_1. In this situation, there is very little precautionary saving by the households and their higher discount rate leads them to low overall saving.

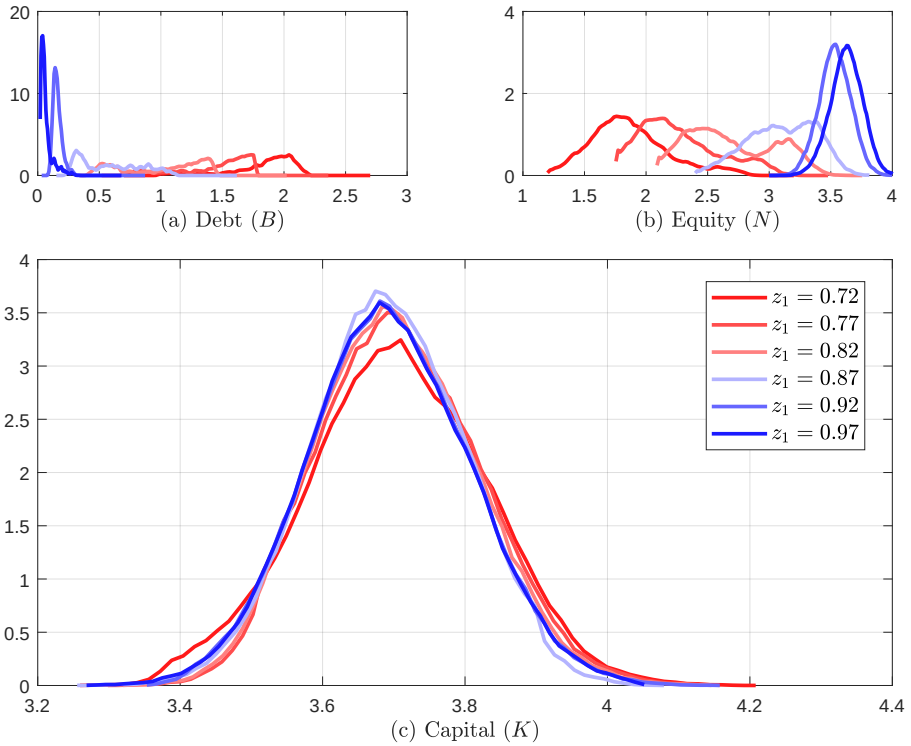

Figure 16: Ergodic distributions of debt, equity, and capital as a function of z_1.

Figure E.4 in Appendix E plots the leverage at the LL-SSS, the unstable SSS, the HL-SSS, and the DSS as a function of z_1. Levels of leverage are decreasing as we raise z_1, reflecting the lower level of precautionary saving.

7.2 The interaction of aggregate and idiosyncratic risk

Figures 11 and 14 documented how the behavior of the model changed as we moved either aggregate or idiosyncratic risk. Figure 17 shows the consequences of moving both aggregate and idiosyncratic risk simultaneously and, as such, it is a good summary of most of our results. Each point in the figure represents a different combination of values of z_1 and σ, with the associated colors of intermediate values computed with a nearest-neighbors algorithm.

There are three regions in the figure. For high levels of idiosyncratic risk (i.e., low z_1), and due to the subsequent high precautionary behavior of households, there is only one HL-SSS. The region becomes larger (i.e., for higher z_1) as we increase σ.

For intermediate values of z_1 (between 0.7 and 0.9) and moderate levels of aggregate risk (σ below 0.02; recall the maximum likelihood point estimate of 0.0142 using U.S. data), we have an HL-SSS and an LL-SSS, exactly as discussed in previous pages. The region is decreasing on σ: as aggregate risk increases, households exhibit more precautionary behavior and the HL-SSS disappears.

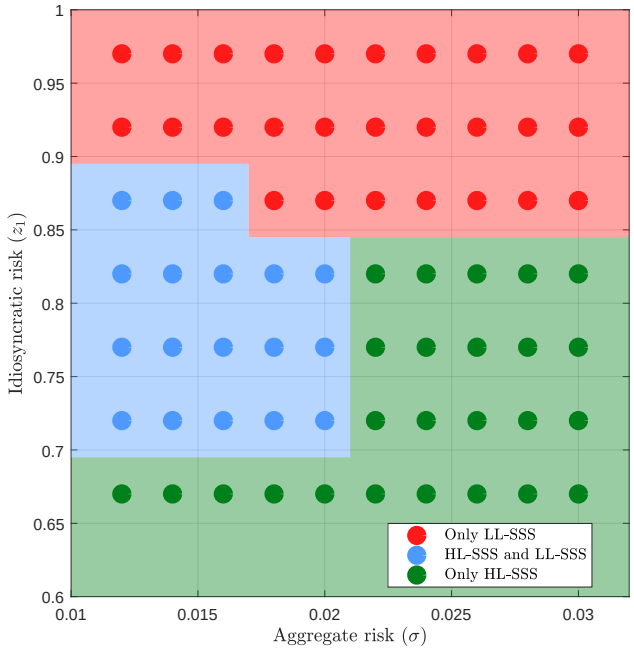

Figure 17: Multiplicity of SSS(s) as a function of aggregate and idiosyncratic volatility.

Finally, for high values of z_1, the level of idiosyncratic risk is sufficiently low that only the LL-SSS exists. One can think about the top row of points, when $z_1 = 0.97$, as a version of the model with very low household heterogeneity. While this is not strictly the representative household version of the model, idiosyncratic risk is so small (low productivity is just 3% below the ergodic mean, and only 5% of households are in such a situation) that we approximate very well the limit case of a representative household.[21]

Figure 17 demonstrates the importance of household heterogeneity in terms of the existence of different SSS(s). If we were going to look at a version of the model with low or no household heterogeneity (the red region), we would only discover a unique SSS and the model would have quite disparate properties (the next subsection will illustrate some of them). Also, such a model would need different parameter values to match the U.S. data, biasing the usefulness of the model for counterfactual analysis and welfare evaluation.

7.3 GIRFs: High vs. low household heterogeneity

Figure 18 reports the GIRFs of the baseline, high household heterogeneity version of the model (continuous green line) to a two-standard-deviations negative capital shock when the economy is at the HL-SSS. These GIRFs are, by construction, identical to the GIRFs (also in continuous

[21] We do not compute the exact representative household version of the model because it would require a different algorithm than the one we use for the heterogeneous case. Consequently, there would be some numerical chatter between both solutions that might complicate the evaluation of differences. Hence, we prefer to set $z_1 = 0.97$ (as high as our algorithm can go before breaking down).

green lines) in Figure 6. Figure 18 also plots the GIRFs at the unique SSS existing when $z_1 = 0.97$, the value in the top row in Figure 17.

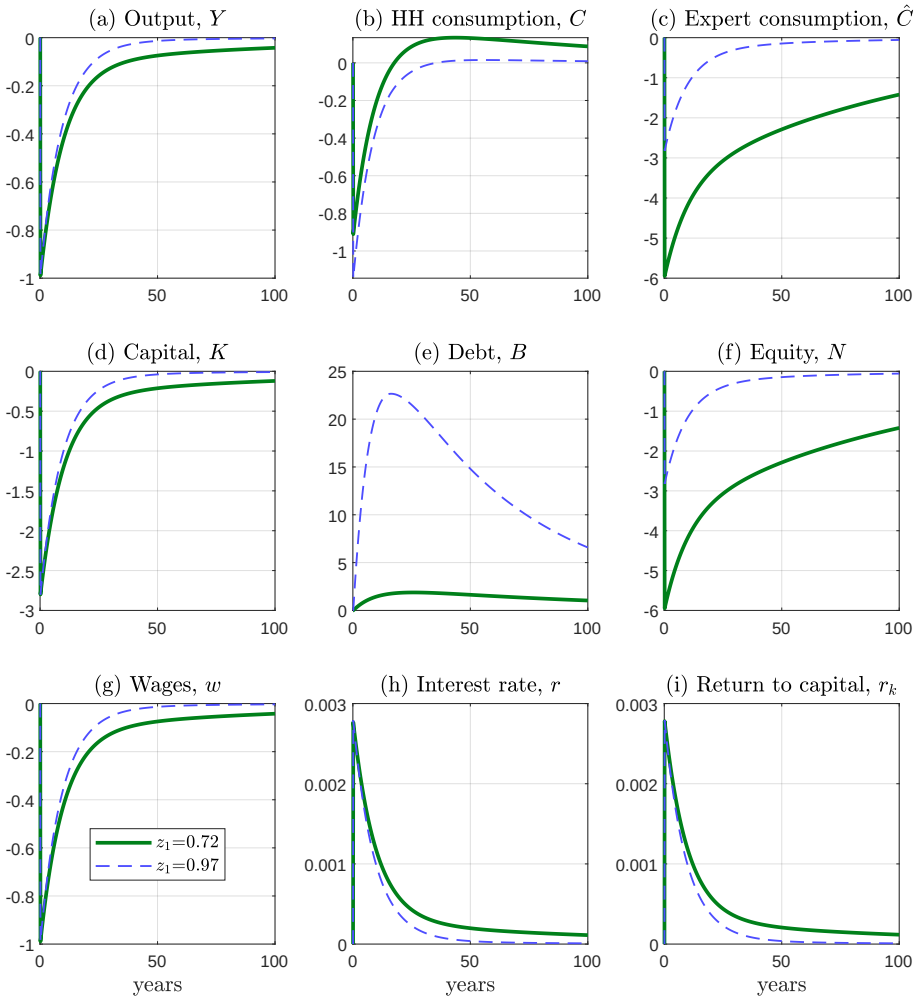

Figure 18: GIRFS, different levels of heterogeneity.

Both sets of GIRFs vary considerably. The fall in output (panel a) is more persistent when we have more heterogeneity. We will show below how this is related to the dynamics of consumption and savings by wealthy households. The reduction in the households' total consumption (panel b) is lower at impact when heterogeneity is high. In comparison, the consumption of the expert drops much more (panel c). This is because the expert starts with higher leverage, and thus its equity drops more (panel f) and it does not issue much additional debt for a long time (panel e).

Figures 14-18 also connect changes in microeconomic conditions with aggregate outcomes. Imagine an economy that, due to technological change or structural transformation, starts having a more turbulent labor market, with households rotating in and out of unemployment more often (or suffering, in an equivalent interpretation of z, more changes to their wages while employed). The increased precautionary saving lowers the risk-free interest rate and increases, on average,

leverage. Thus, the economy becomes more volatile even when the volatility of the aggregate shocks remains constant.

This result shows how our model simultaneously accounts for more wealth inequality among capital owners before the financial crisis, more debt, lower risk-free rates, and a heightened fragility of the economy to negative capital shocks.

7.4 Inspecting the heterogeneity mechanism

The key to understanding the differences in Figure 18 is to explore how the reduction in households' total consumption (panel b) is distributed among households. After a negative aggregate shock, wages decrease and the risk-free rate increases. Poorer households, mainly dependent on labor earnings, lose much. In comparison, richer households, with more income from their assets, lose less or, if they are sufficiently wealthy, can even gain (see, in the third row of Figure F.7 in Appendix F, how the value function improves for households with high assets, in particular in the HL-SSS).

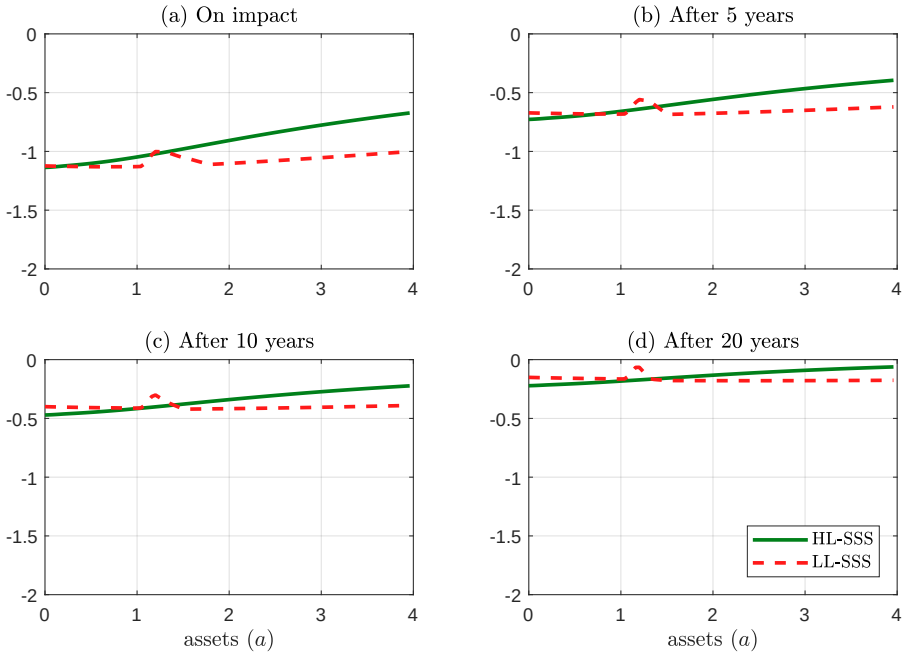

Figure 19: Difference in consumption decision rules at different points in time after the shock.

Even more, rich households can take advantage of a higher risk-free rate to accumulate wealth fast. They reduce their consumption not so much as a consequence of lower labor income (wealth effect), but as a consequence of better rewards for savings (substitution effect). Moreover, the sluggish aggregate dynamics of the model mean that wealthy households expect a higher risk-free rate to last for a long time. Because of the marginal decreasing utility of consumption, the integral of these consumption responses across households is different than the response of a

household with average wealth. Hence, the responses of total consumption and the evolution of total debt are also different.

Figure 19 documents the heterogeneity of consumption effects. The figure plots the consumption decision rules for high-productivity households ($z = z_2$) along the asset axis (results for $z = z_1$ are qualitatively similar) when a two-standard-deviations negative capital shock hits the economy at the HL-SSS (continuous green line). To facilitate interpretation, we plot the *difference* in the consumption decision rules with respect to the case without the shock at different points in time: at impact (panel a); after 5 years (panel b); after 10 years (panel c); and after 20 years (panel d).

At impact, all households reduce their consumption, but poorer households do so by a larger amount, reflecting their lower income. For instance, richer households (level of assets of 4) reduce their consumption around one-third less than poor households (level of assets of 0). The difference in consumption reduction survives over time.

The asymmetry in the consumption response is much smaller when we have a two-standard-deviations negative capital shock at the LL-SSS. Since the risk-free interest rate is less persistent in this case, the intertemporal substitution mechanism is weaker.

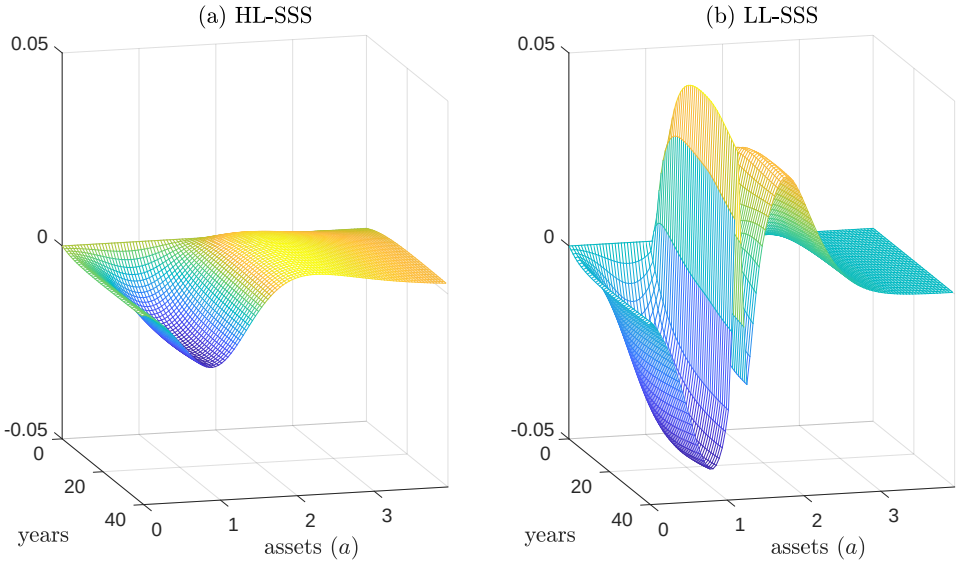

Figure 20: DIRFs at the HL-SSS and LL-SSS.

The asymmetric consumption responses have a direct impact on how the wealth distribution evolves. To illustrate this point, Figure 20 draws what we call the *distributional impulse response functions*, or DIRFs. A DIRF is the natural analog of a GIRF except that, instead of plotting the evolution of an aggregate variable such as output or wages, we plot the evolution of the wealth density $g_t(\cdot)$. More concretely, Figure 20 plots the difference between the density before and after the shock, $g_t(\cdot) - g_0(\cdot)$. Time, in years, is plotted on the y-axis, assets on the x-axis, and the

DIRFs on the z-axis. A positive value of the DIRF at a given asset level and point in time should be read as the density is higher at that asset level and point in time than it would have been in the absence of a shock. A negative value has the opposite interpretation. In the left panel of Figure 20, we plot the DIRF to a two-standard-deviations negative capital shock when the economy is at the HL-SSS. In the right panel, we plot the DIRF to a two-standard-deviations negative capital shock when the economy is at the LL-SSS.

In panel (a), we can see how households with low assets must draw from their wealth to smooth consumption (even if consumption still drops) to compensate for lower income. This mechanism makes the DIRF negative in that region. In comparison, households with higher assets are reducing their consumption to respond to a temporarily higher risk-free rate and accumulate wealth. Thus, the DIRF is positive in the region of high assets. These effects are more pronounced in panel (b). In the LL-SSS, poor households have too little debt to smooth consumption, and wealthy households accumulate much additional debt as the risk-free interest rate changes.

7.5 On the approximation of the PLM

The previous four subsections summarize why household heterogeneity matters in this economy. But then, why can we approximate our PLM well with just one moment of the distribution? Because the consumption decision rule of the households is close to linear with respect to the household state variables.[22] Interestingly, it is so despite the fact that it is strongly nonlinear with respect to the aggregate state variables. As a function of (B, N), the consumption decision rule of the households is strongly nonlinear, leading to the multiplicity of SSS(s) and endogenous regime-switching we have discussed. Households understand the nonlinear responses of wages and the risk-free rate they receive from labor and savings with respect to (B, N) and behave accordingly. See the different decision rules in Figure 21.

This point is much more general than its operation in our model: we can have models with agents' heterogeneity that imply nearly linear decision rules of the agents with respect to individual state variables and very nonlinear decision rules with respect to aggregate state variables. In that case, we will need solution methods that can capture such a nonlinearity and machine learning provides the required flexibility.

[22] Only agents close to the borrowing limit face a nonlinear consumption decision rule but, being close to zero assets, they contribute relatively little to the aggregate dynamics of capital. And, as shown in Figure 13, there are not that many of them.

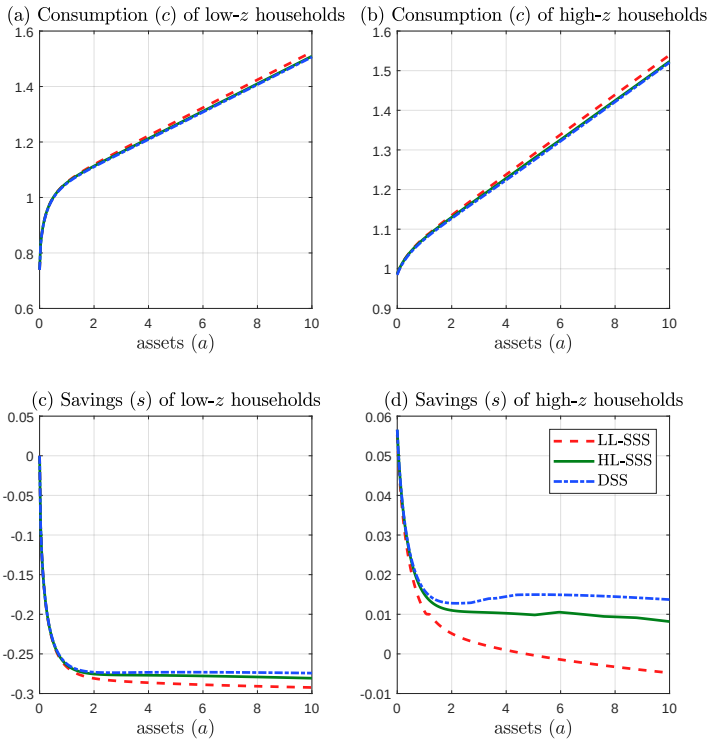

Figure 21: Consumption and saving functions.

8 Conclusion

In this paper, we have postulated, computed, and estimated a continuous-time DSGE model with a financial sector and heterogeneous households. This exercise has allowed us to argue that the wealth distribution among households is central to the analysis of the effects of financial frictions. More concretely, we have shown how the total level of endogenous aggregate risk creates an endogenous regime-switching process for output, interest rates, debt, and leverage. The regime-switching generates i) multimodal distributions of the variables above; ii) time-varying levels of volatility and skewness for the same variables; and iii) supercycles of borrowing and deleveraging.

Methodologically, this paper also shows how to compute and estimate a continuous-time model with heterogeneous agents efficiently. For the computation, we have exploited tools borrowed from machine learning. For the estimation, we have built on contributions from inference with diffusions. Our methodology can be useful to analyze other heterogeneous agents models with aggregate shocks that display significant nonlinear features. An obvious candidate is the heterogeneous agent New Keynesian (HANK) model with a zero lower bound (ZLB) on the nominal interest rates, such as Auclert (2016), Auclert and Rognlie (2018), Gornemann et al. (2012), Kaplan et al. (2018), Luetticke (2015), and McKay et al. (2016). The ZLB introduces a nonlinearity in the state space of aggregate variables that cannot be addressed either with local methods or with global methods based on linear laws of motion.

References

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2017). Income and wealth distribution in macroeconomics: A continuous-time approach. Working Paper 23732, National Bureau of Economic Research.

Adrian, T. and Boyarchenko, N. (2012). Intermediary leverage cycles and financial stability. Staff Report 567, Federal Reserve Bank of New York.

Adrian, T., Boyarchenko, N., and Giannone, D. (2019a). Multimodality in macro-financial dynamics. Staff Report 903, Federal Reserve Bank of New York.

Adrian, T., Boyarchenko, N., and Giannone, D. (2019b). Vulnerable growth. *American Economic Review*, 109(4):1263–89.

Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *Journal of Financial Intermediation*, 19(3):418–437.

Ahn, S., Kaplan, G., Moll, B., Winberry, T., and Wolf, C. (2017). When inequality matters for macro and macro matters for inequality. In *NBER Macroeconomics Annual 2017, Vol. 32*. University of Chicago Press.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, 109(3):659–684.

Algan, Y., Allais, O., and Den Haan, W. J. (2008). Solving heterogeneous-agent models with parameterized cross-sectional distributions. *Journal of Economic Dynamics and Control*, 32(3):875–908.

Algan, Y., Allais, O., Den Haan, W. J., and Rendahl, P. (2014). Solving and simulating models with heterogeneous agents and aggregate uncertainty. In Schmedders, K. and Judd, K. L., editors, *Handbook of Computational Economics*, volume 3, pages 277–324. Elsevier.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2017). Global inequality dynamics: New findings from wid.world. Working Paper 23119, National Bureau of Economic Research.

Andersen, T. G. and Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77(2):343–377.

Andreasen, M., Fernández-Villaverde, J., and Rubio-Ramírez, J. (2018). The pruned state-space system for non-linear DSGE models: Theory and empirical applications. *Review of Economic Studies*, 85:1–49.

Auclert, A. (2016). Monetary policy and the redistribution channel. Stanford University.

Auclert, A., Bardóczy, B., Rognlie, M., and Straub, L. (2019). Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models. Mimeo, Stanford University.

Auclert, A. and Rognlie, M. (2018). Inequality and aggregate demand. Stanford University.

Azinović, M., Gaegauf, L., and Scheidegger, S. (2020). Deep equilibrium nets. University of Lausanne.

Bach, F. (2017). Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18(1):629–681.

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945.

Basak, S. and Cuoco, D. (1998). An equilibrium model with restricted stock market participation. *Review of Financial Studies*, 11(2):309–341.

Bayer, C. and Luetticke, R. (2018). Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods. Mimeo, University of Bonn.

Bergin, J. and Bernhardt, D. (1992). Anonymous sequential games with aggregate uncertainty. *Journal of Mathematical Economics*, 21(6):543–562.

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1, pages 1341–1393. Elsevier.

Biggio, S. and Sannikov, Y. (2019). A model of intermediation, money, interest, and prices. UCLA.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag.

Blanchard, O. (1997). The Medium Run. *Brookings Papers on Economic Activity*, 28(2):89–158.

Blanchard, O. and Galí, J. (2010). Labor markets and monetary policy: A New Keynesian model with unemployment. *American Economic Journal: Macroeconomics*, 2(2):1–30.

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421.

Candler, G. V. (1999). Finite difference methods for continuous time dynamic programming. In Marimón, R. and Scott, A., editors, *Computational Methods for the Study of Dynamic Economies*, pages 172–194. Oxford University Press.

Chacko, G. and Viceira, L. M. (2003). Spectral GMM estimation of continuous-time processes. *Journal of Econometrics*, 116(1):259–292.

Cho, I.-K. (1995). Perceptrons play the repeated prisoner's dilemma. *Journal of Economic Theory*, 67(1):266–284.

Cho, I.-K. and Sargent, T. J. (1996). Neural networks for encoding and adapting in dynamic economies. In Amman, H., Kendrick, D., and Rust, J., editors, *Handbook of Computational Economics*, volume 1, pages 441–470. Elsevier.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314.

Den Haan, W. J. (1996). Heterogeneity, aggregate uncertainty, and the short-term interest rate. *Journal of Business & Economic Statistics*, 14(4):399–411.

Den Haan, W. J. (1997). Solving dynamic models with aggregate shocks and heterogeneous agents. *Macroeconomic Dynamics*, 1(02):355–386.

Den Haan, W. J. and Rendahl, P. (2010). Solving the incomplete markets model with aggregate uncertainty using explicit aggregation. *Journal of Economic Dynamics and Control*, 34(1):69–78.

Dew-Becker, I., Tahbaz-Salehi, A., and Vedolin, A. (2019). Macro skewness and conditional second moments: evidence and theories. Technical report, Northwestern University.

Drautzburg, T., Fernández-Villaverde, J., and Guerrón-Quintana, P. (2017). Political distribution risk and aggregate fluctuations. Working Paper 23647, National Bureau of Economic Research.

Duarte, V. (2018). Sectoral reallocation and endogenous risk-aversion: Solving macro-finance models with machine learning. MIT Sloan School of Management.

Duffie, D. and Epstein, L. (1992). Stochastic differential utility. *Econometrica*, 60(2):353–94.

Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press.

Fernández-Villaverde, J. and Guerrón-Quintana, P. (2020). Uncertainty shocks and business cycle research. NBER Working Papers 26768, National Bureau of Economic Research, Inc.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies*, 74(4):1059–1087.

Galí, J. and Gambetti, L. (2009). On the sources of the Great Moderation. *American Economic Journal: Macroeconomics*, 1(1):26–57.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Gornemann, N., Kuester, K., and Nakajima, M. (2012). Monetary policy with heterogeneous agents. Working Paper 12-21, Federal Reserve Bank of Philadelphia.

Greenwood, J., Hercowitz, Z., and Krusell, P. (1997). Long-run implications of investment-specific technological change. *American Economic Review*, 87(3):342–62.

Hall, R. E. and Milgrom, P. R. (2008). The limited influence of unemployment on the wage bargain. *American Economic Review*, 98(4):1653–74.

He, Z. and Krishnamurthy, A. (2012). A model of capital and crises. *Review of Economic Studies*, 79(2):735–777.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–70.

Holston, K., Laubach, T., and Williams, J. C. (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, 108:S59 – S75.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5-6):953–969.

Jordà, O., Schularick, M., and Taylor, A. M. (2016). *Macrofinancial History and the New Business Cycle Facts*, pages 213–263. University of Chicago Press.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to HANK. *American Economic Review*, 108(3):697–743.

Krusell, P. and Smith, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896.

Kubler, F. and Scheidegger, S. (2018). Self-justified equilibria: Existence and computation. Working Paper, DBF, University of Zurich.

Kumhof, M., Ranciére, R., and Winant, P. (2015). Inequality, leverage, and crises. *American Economic Review*, 105(3):1217–45.

Liu, P., Theodoridis, K., Mumtaz, H., and Zanetti, F. (2018). Changing macroeconomic dynamics at the zero lower bound. *Journal of Business & Economic Statistics*, 0(0):1–14.

Lo, A. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory*, 4(2):231–247.

Luetticke, R. (2015). Transmission of monetary policy and heterogeneity in household portfolios. UCL.

Maliar, L., Maliar, S., and Valli, F. (2010). Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm. *Journal of Economic Dynamics and Control*, 34(1):42–49.

Maliar, L., Maliar, S., and Winant, P. (2019). Will artificial intelligence replace computational economists any time soon? CEPR Discussion Papers DP14024, CEPR.

McKay, A., Nakamura, E., and Steinsson, J. (2016). The power of forward guidance revisited. *American Economic Review*, 106(10):3133–58.

McKinsey Global Institute (2010). Debt and deleveraging: The global credit bubble and its economic consequences. Technical report, McKinsey Global Institute.

Miao, J. (2006). Competitive equilibria of economies with a continuum of consumers and aggregate shocks. *Journal of Economic Theory*, 128(1):274–298.

Moffitt, R. and Zhang, S. (2018). Income volatility and the PSID: Past research and new results. *AEA Papers and Proceedings*, 108:277–80.

Nuño, G. and Thomas, C. (2016). Optimal monetary policy with heterogeneous agents. Working Paper 1624, Banco de España.

Nuño, G. and Thomas, C. (2017). Bank leverage cycles. *American Economic Journal: Macroeconomics*, 9(2):32–72.

Pröhl, E. (2015). Approximating equilibria with ex-post heterogeneity and aggregate risk. Research Paper 17-63, Swiss Finance Institute.

Reinhart, C. M. and Rogoff, K. S. (2009). *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, 33(3):649–665.

Reiter, M. (2010). Solving the incomplete markets model with aggregate uncertainty by backward induction. *Journal of Economic Dynamics and Control*, 34(1):28–35.

Ríos-Rull, J.-V. (1997). Computation of equilibria in heterogeneous agent models. Staff Report 231, Federal Reserve Bank of Minneapolis.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323:533–536.

Sager, E. (2014). Solving the incomplete markets model with aggregate uncertainty: The method of mixtures. Bureau of Labor Statistics, Price Research Division.

Scheidegger, S. and Bilionis, I. (2017). Machine learning for high-dimensional dynamic stochastic economies. UCL.

Winberry, T. (2018). A method for solving and estimating heterogeneous agent macro models. *Quantitative Economics*, 9(3):1123–1151.

Appendix

This appendix compiles further details about the equilibrium conditions, the numerical aspects of our solution method, and additional results not reported in the main text.

A Equilibrium conditions

We can stack all the equilibrium conditions of the model (except the optimality condition for households) in two blocks. The first block includes all the variables that depend directly on N_t, B_t, and dZ_t:

$$w_t = (1-\alpha) K_t^{\alpha} \tag{30}$$

$$rc_t = \alpha K_t^{\alpha-1} \tag{31}$$

$$r_t = \alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t} \tag{32}$$

$$dr_t^k = (rc_t - \delta)\, dt + \sigma dZ_t \tag{33}$$

$$dN_t = \left(\alpha K_t^{\alpha-1} - \delta - \widehat{\rho} - \sigma^2 \left(1 - \frac{K_t}{N_t}\right) \frac{K_t}{N_t}\right) N_t dt + \sigma K_t dZ_t. \tag{34}$$

The second block includes the equations determining the aggregate consumption of the households, dB_t, dK_t, and $\frac{\partial g_{it}}{\partial t}$:

$$C_t \equiv \sum_{i=1}^{2} \int c\left(a, z, K_t, G_t\right) g_{it}\left(a\right) da \tag{35}$$

$$dB_t = \left(\left(1-\alpha\right) K_t^{\alpha} + \left(\alpha K_t^{\alpha-1} - \delta - \sigma^2 \frac{K_t}{N_t}\right) B_t - C_t\right) dt \tag{36}$$

$$dK_t = dN_t + dB_t \tag{37}$$

$$\frac{\partial g_{it}}{\partial t} = -\frac{\partial}{\partial a}\left(s\left(a, z, K_t, G_t\right) g_{it}(a)\right) - \lambda_i g_{it}(a) + \lambda_j g_{jt}(a),\ i \neq j = 1, 2. \tag{38}$$

The second block shows i) how the density $\left\{g_{it}\left(\cdot\right)\right\}_{t\geq 0}$ for $i \in \{1,2\}$ matters to determine C_t, ii) that C_t pins down dB_t, and iii) that once we have dB_t, we can calculate dK_t. Hence, computing the equilibrium of this economy is equivalent to finding C_t. Once C_t is known, all other aggregate variables follow directly.

B Numerical algorithm

We describe the numerical algorithm used to solve for the equilibrium value function, $v(a, z, B, N)$, the density $g(a, z, B, N)$, and the aggregate debt B and equity N. The algorithm proceeds in 3 steps. We describe each step in turn.

Step 1: Solution to the Hamilton-Jacobi-Bellman equation

The HJB equation is solved using an *upwind finite difference* scheme similar to Candler (1999) and Achdou et al. (2017). It approximates the value function $v_i(a, B, N)$, $i = 1, 2$ on a finite grid with steps Δa, ΔB, $\Delta N : a \in \{a_1, ..., a_J\}, B \in \{B_1, ..., B_L\}, N \in \{N_1, ..., N_M\}$, where:

$$\begin{aligned} a_j &= a_{j-1} + \Delta a = a_1 + (j-1)\Delta a, \ 2 \leq j \leq J, \\ B_l &= B_{l-1} + \Delta B = B_1 + (l-1)\Delta L, \ 2 \leq l \leq L, \\ N_m &= N_{m-1} + \Delta N = N_1 + (m-1)\Delta N, \ 2 \leq m \leq M. \end{aligned}$$

The lower bound in the wealth space is $a_1 = 0$, such that $\Delta a = a_J/(J-1)$. We use the notation $v_{i,j,l,m} \equiv v_i(a_j, B_l, N_m)$, and similarly for the policy function $c_{i,j,l,m}$. The derivatives are evaluated according to

$$\begin{aligned} \frac{\partial_i v(a_j, B_l, N_m)}{\partial a} &\approx \partial_f v_{i,j,l,m} \equiv \frac{v_{i,j+1,l,m} - v_{i,j,l,m}}{\Delta a}, \\ \frac{\partial_i v(a_j, B_l, N_m)}{\partial a} &\approx \partial_b v_{i,j,l,m} \equiv \frac{v_{i,j,l,m} - v_{i,j-1,l,m}}{\Delta a}. \\ \frac{\partial_i v(a_j, B_l, N_m)}{\partial B} &\approx \partial_B v_{i,j,l,m} \equiv \frac{v_{i,j,l+1,m} - v_{i,j,l,m}}{\Delta B}, \\ \frac{\partial_i v(a_j, B_l, N_m)}{\partial Z} &\approx \partial_N v_{i,j,l,m} \equiv \frac{v_{i,j,l,m+1} - v_{i,j,l,m}}{\Delta N}, \\ \frac{\partial_i^2 v(a_j, B_l, N_m)}{\partial N^2} &\approx \partial_{NN}^2 v_{i,j,l,m} \equiv \frac{v_{i,j,l,m+1} + v_{i,j,l,m-1} - 2v_{i,j,l,m}}{(\Delta N)^2}. \end{aligned}$$

At each point of the grid, the first derivative with respect to a can be approximated with a forward (f) or a backward (b) approximation. In an upwind scheme, the choice of forward or backward derivative depends on the sign of the *drift function* for the state variable, given by:

$$s_{i,j,l,m} \equiv w_{l,m} z_i + r_{l,m} a_j - c_{i,j,l,m}, \tag{39}$$

where:

$$
\begin{aligned}
c_{i,j,l,m} &= \left[\frac{\partial v_{i,j,l,m}}{\partial a}\right]^{-1/\gamma}, && (40)\\
w_{l,m} &= (1-\alpha)\, Z\left(B_l + N_m\right)^{\alpha}, && (41)\\
r_{l,m} &= \alpha Z\left(B_l + N_m\right)^{\alpha-1} - \delta - \sigma^2 \frac{\left(B_l + N_m\right)}{N_m}. && (42)
\end{aligned}
$$

Let superscript n denote the iteration counter. The HJB equation is approximated by the following upwind scheme:

$$
\begin{aligned}
\frac{v^{n+1}_{i,j,l,m} - v^{n}_{i,j,l,m}}{\Delta} + \rho v^{n+1}_{i,j,l,m} &= \frac{\left(c^{n}_{i,j,l,m}\right)^{1-\gamma} - 1}{1-\gamma} + \partial_f v^{n+1}_{i,j,l,m} s^{n}_{i,j,l,m,f} \mathbf{1}_{s^{n}_{i,j,n,m,f}>0} + \partial_B v^{n+1}_{i,j,l,m} s^{n}_{i,j,l,m,b} \mathbf{1}_{s^{n}_{i,j,l,m,b}<0}\\
&\quad + \lambda_i \left(v^{n+1}_{-i,j,l,m} - v^{n+1}_{i,j,l,m}\right) + h_{l,m}\partial_B v_{i,j,l,m} + \mu^N_{l,m}\partial_N v_{i,j,l,m}\\
&\quad + \frac{\left[\sigma^N_{l,m}\right]^2}{2}\partial^2_{NN} v_{i,j,l,m}
\end{aligned}
$$

for $i = 1, 2$, $j = 1, ..., J$, $l = 1, .., L$, $m = 1, ..., M$, where $\mathbf{1}(\cdot)$ is the indicator function and

$$
\begin{aligned}
h_{l,m} &\equiv h\left(B_l, N_m\right),\\
\mu^N_{l,m} &\equiv \mu^N\left(B_l, N_m\right) = \alpha Z\left(B_l + N_m\right)^{\alpha} - \delta\left(B_l + N_m\right) - r_{l,m} B_l - \widehat{\rho} N_m,\\
\sigma^N_{l,m} &\equiv \sigma^N\left(B_l, N_m\right) = \sigma\left(B_l + N_m\right),\\
s^{n}_{i,j,l,m,f} &= w_{l,m} z_i + r_{l,m} a_j - \left[\frac{1}{\partial^n_f v_{i,j,l,m}}\right]^{1/\gamma},\\
s^{n}_{i,j,l,m,b} &= w_{l,m} z_i + r_{l,m} a_j - \left[\frac{1}{\partial^n_b v_{i,j,l,m}}\right]^{1/\gamma}.
\end{aligned}
$$

Thus, when the drift is positive ($s^{n}_{i,j,l,m,f} > 0$), we employ a forward approximation of the derivative, $\partial^n_f v_{i,j,l,m}$; when it is negative ($s^{n}_{i,j,l,m,b} < 0$), we employ a backward approximation, $\partial^n_b v_{i,j,l,m}$. The term $\frac{v^{n+1}_{i,j,l,m} - v^{n}_{i,j,l,m}}{\Delta} \to 0$ as $v^{n+1}_{i,j,l,m} \to v^{n}_{i,j,l,m}$.

Moving all terms involving v^{n+1} to the left-hand side and the rest to the right-hand side, we obtain:

$$
\begin{aligned}
\frac{v^{n+1}_{i,j,l,m} - v^{n}_{i,j,l,m}}{\Delta} + \rho v^{n+1}_{i,j,l,m} &= \frac{\left(c^{n}_{i,j,n,m}\right)^{1-\gamma} - 1}{1-\gamma} + v^{n+1}_{i,j-1,l,m}\alpha^{n}_{i,j,l,m} + v^{n+1}_{i,j,l,m}\beta^{n}_{i,j,l,m} + v^{n+1}_{i,j+1,l,m}\xi^{n}_{i,j,l,m}\\
&\quad + \lambda_i v^{n+1}_{-i,j,l,m} + v^{n+1}_{i,j,l+1,m}\frac{h_{l,m}}{\Delta B} + v^{n+1}_{i,j,l,m+1}\varkappa_{l,m} + v^{n+1}_{i,j,l,m-1}\varrho_{l,m} \qquad (43)
\end{aligned}
$$

where:

$$\begin{aligned}
\alpha_{i,j}^{n} &\equiv -\frac{s_{i,j,B}^{n}\mathbf{1}_{s_{i,j,B}^{n}<0}}{\Delta a},\\
\beta_{i,j,l,m}^{n} &\equiv -\frac{s_{i,j,l,m,f}^{n}\mathbf{1}_{s_{i,j,n,mF}^{n}>0}}{\Delta a}+\frac{s_{i,j,l,m,b}^{n}\mathbf{1}_{s_{i,j,l,m,b}^{n}<0}}{\Delta a}-\lambda_{i}-\frac{h_{l,m}}{\Delta B}-\frac{\mu_{l,m}^{N}}{\Delta N}-\frac{\left(\sigma_{l,m}^{N}\right)^{2}}{\left(\Delta N\right)^{2}},\\
\xi_{i,j}^{n} &\equiv \frac{s_{i,j,F}^{n}\mathbf{1}_{s_{i,j,F}^{n}>0}}{\Delta a},\\
\varkappa_{l,m} &\equiv \frac{\mu_{l,m}^{N}}{\Delta N}+\frac{\left(\sigma_{l,m}^{N}\right)^{2}}{2\left(\Delta N\right)^{2}}=\frac{\left[\alpha Z\left(B_{l}+N_{m}\right)^{\alpha}-\delta\left(B_{l}+N_{m}\right)-r_{l,m}B_{l}-\widehat{\rho}N_{m}\right]}{\Delta N}+\frac{\sigma^{2}\left(B_{l}+N_{m}\right)^{2}}{2\left(\Delta N\right)^{2}},\\
\varrho_{l,m} &\equiv \frac{\left(\sigma_{l,m}^{N}\right)^{2}}{2\left(\Delta N\right)^{2}}=\frac{\sigma^{2}\left(B_{l}+N_{m}\right)^{2}}{2\left(\Delta N\right)^{2}}.
\end{aligned}$$

for $i=1,2$, $j=1,...,J$, $l=1,..,L$, $m=1,...,M$. We consider boundary state constraints in a ($s_{i,1,B}^{n}=s_{i,J,F}^{n}=0$). The boundary conditions in B and N are reflections.

In equation (43), the optimal consumption is set to:

$$c_{i,j,n,m}^{n}=\left(\partial v_{i,j,l,m}^{n}\right)^{-1/\gamma}. \tag{44}$$

where:

$$\partial v_{i,j,l,m}^{n}=\partial_{f}v_{i,j,l,m}^{n}\mathbf{1}_{s_{i,j,n,mF}^{n}>0}+\partial_{b}v_{i,j,l,m}^{n}\mathbf{1}_{s_{i,j,l,m,b}^{n}<0}+\partial\bar{v}_{i,j,l,m}^{n}\mathbf{1}_{s_{i,j,n,mF}^{n}\leq 0}\mathbf{1}_{s_{i,j,l,m,b}^{n}\geq 0}.$$

In the above expression, $\partial\bar{v}_{i,j,l,m}^{n}=(\bar{c}_{i,j,n,m}^{n})^{-\gamma}$ where $\bar{c}_{i,j,n,m}^{n}$ is the consumption level such that the drift is zero:

$$\bar{c}_{i,j}^{n}=w_{l,m}z_{i}+r_{l,m}a_{j}.$$

We define:

$$\mathbf{A}^n_{l.m} = \begin{bmatrix} \beta^n_{1,1,l,m} & \xi^n_{1,1,l,m} & 0 & 0 & \cdots & 0 & \lambda_1 & 0 & \cdots & 0 \\ \alpha^n_{1,2,l,m} & \beta^n_{1,2,l,m} & \xi^n_{1,2,l,m} & 0 & \cdots & 0 & 0 & \lambda_1 & \ddots & 0 \\ 0 & \alpha^n_{1,3,l,m} & \beta^n_{1,3,l,m} & \xi^n_{1,3,l,m} & \cdots & 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha^n_{1,J-1,l,m} & \beta^n_{1,J-1,l,m} & \xi^n_{1,J-1,l,m} & 0 & \cdots & \lambda_1 & 0 \\ 0 & 0 & \cdots & 0 & \alpha^n_{1,J,l,m} & \beta^n_{1,J,l,m} & 0 & 0 & \cdots & \lambda_1 \\ \lambda_2 & 0 & \cdots & 0 & 0 & 0 & \beta^n_{2,1,l,m} & \xi^n_{2,1,l,m} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \lambda_2 & 0 & \cdots & \alpha^n_{2,J,l,m} & \beta^n_{2,J,l,m} \end{bmatrix},$$

$$\mathbf{v}^{n+1}_{l,m} = \begin{bmatrix} \mathbf{v}^{n+1}_{1,1,l,m} \\ \mathbf{v}^{n+1}_{1,2,l,m} \\ \vdots \\ \mathbf{v}^{n+1}_{1,J,l,m} \\ \mathbf{v}^{n+1}_{2,1,l,m} \\ \vdots \\ \mathbf{v}^{n+1}_{2,J,l,m} \end{bmatrix}$$

and

$$\mathbf{A}^n_m = \begin{bmatrix} \mathbf{A}^n_{1,m} & \frac{h_{1,m}}{\Delta K}\mathbf{I}_{2J} & \mathbf{0}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \mathbf{A}^n_{2,m} & \frac{h_{2,m}}{\Delta K}\mathbf{I}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{A}^n_{3,m} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & & \mathbf{0}_{2J} & \mathbf{A}^n_{L-1,m} & \frac{h_{L-1,m}}{\Delta K}\mathbf{I}_{2J} \\ \mathbf{0}_{2J} & \mathbf{0}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \left(\mathbf{A}^n_{L,m} + \frac{h_{L,m}}{\Delta K}\mathbf{I}_{2J}\right) \end{bmatrix}, \quad \mathbf{v}^{n+1}_m = \begin{bmatrix} \mathbf{v}^{n+1}_{1,m} \\ \mathbf{v}^{n+1}_{2,m} \\ \vdots \\ \mathbf{v}^{n+1}_{L,m} \end{bmatrix},$$

where $\mathbf{I}_n$ and $\mathbf{0}_n$ are the identity matrix and the zero matrix of dimension $n \times n$, respectively.

We can also define

$$\mathbf{A}^n = \begin{bmatrix} (\mathbf{A}_1^n + \mathbf{P}_1) & \mathbf{X}_1 & \mathbf{0}_{2J\times L} & \cdots & \mathbf{0}_{2J\times L} & \mathbf{0}_{2J\times L} \\ \mathbf{P}_2 & \mathbf{A}_2^n & \mathbf{X}_2 & \cdots & \mathbf{0}_{2J\times L} & \mathbf{0}_{2J\times L} \\ \mathbf{0}_{2J\times L} & \mathbf{P}_3 & \mathbf{A}_3^n & \cdots & \mathbf{0}_{2J\times L} & \mathbf{0}_{2J\times L} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & & \mathbf{P}_{M-1} & \mathbf{A}_{M-1}^n & \mathbf{X}_{M-1} \\ \mathbf{0}_{2J\times L} & \mathbf{0}_{2J\times L} & \cdots & \mathbf{0}_{2J\times L} & \mathbf{P}_M & (\mathbf{A}_M^n + \mathbf{X}_M) \end{bmatrix}, \ \mathbf{v}^{n+1} = \begin{bmatrix} \mathbf{v}_1^{n+1} \\ \mathbf{v}_2^{n+1} \\ \vdots \\ \mathbf{v}_M^{n+1} \end{bmatrix},$$

$$\mathbf{X}_m = \begin{bmatrix} \varkappa_{1,m}\mathbf{I}_{2J} & \mathbf{0}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \varkappa_{2,m}\mathbf{I}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & \mathbf{0}_{2J} & \varkappa_{L-1,m}\mathbf{I}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \varkappa_{L,m}\mathbf{I}_{2J} \end{bmatrix},$$

$$\mathbf{P}_m = \begin{bmatrix} \varrho_{1,m}\mathbf{I}_{2J} & \mathbf{0}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \varrho_{2,m}\mathbf{I}_{2J} & \cdots & \mathbf{0}_{2J} & \mathbf{0}_{2J} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & & \mathbf{0}_{2J} & \varrho_{L-1,m}\mathbf{I}_{2J} & \mathbf{0}_{2J} \\ \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \mathbf{0}_{2J} & \varrho_{L,m}\mathbf{I}_{2J} \end{bmatrix}, \mathbf{u}^n = \begin{bmatrix} \frac{(c_{1,1,1,1}^n)^{1-\gamma}-1}{1-\gamma} \\ \frac{(c_{1,2,1,1}^n)^{1-\gamma}-1}{1-\gamma} \\ \vdots \\ \vdots \\ \frac{(c_{2,J,L,M}^n)^{1-\gamma}-1}{1-\gamma} \end{bmatrix}.$$

Then, equation (43) is a system of $2 \times J \times L \times M$ linear equations that can be written in matrix notation as:

$$\frac{1}{\Delta}\left(\mathbf{v}^{n+1} - \mathbf{v}^n\right) + \rho\mathbf{v}^{n+1} = \mathbf{u}^n + \mathbf{A}^n\mathbf{v}^{n+1}.$$

The system in turn can be written as

$$\mathbf{B}^n\mathbf{v}^{n+1} = \mathbf{d}^n \tag{45}$$

where $\mathbf{B}^n = \left(\frac{1}{\Delta} + \rho\right)\mathbf{I} - \mathbf{A}^n$ and $\mathbf{d}^n = \mathbf{u}^n + \frac{1}{\Delta}\mathbf{v}^n$.

The algorithm to solve the HJB equation begins with an initial guess $v_{i,j,l,m}^0$. Set $n = 0$. Then:

1. Compute $c_{i,j,l,m}^n$, $i = 1, 2$ using (44).
2. Find $v_{i,j,l,m}^{n+1}$ solving the linear system of equations (45).
3. If $v_{i,j,l,m}^{n+1}$ is close enough to $v_{i,j,l,m}^n$, stop. If not, set $n := n + 1$ and proceed to step 1.

Most programming languages, such as `Julia` or `Matlab`, include efficient routines to handle sparse matrices such as $\mathbf{A}^n$.

Step 2: Solution to the KF equation

The income-wealth distribution conditional on the current realization of aggregate debt $B = B_l$ and equity $N = N_m$ can be characterized by the KF equation:

$$\frac{\partial g}{\partial t} = -\frac{\partial}{\partial a}\left[s_i\left(a, B, N\right) g_{i,t}(a)\right] - \lambda_i g_{i,t}(a) + \lambda_{-i} g_{-i,t}(a),\ i = 1, 2. \tag{46}$$

$$1 = \int_0^\infty g(a) da. \tag{47}$$

If we define a time step Δt, we also solve this equation using a finite difference scheme. We use the notation $g_{i,j} \equiv g_i(a_j)$. The system can now be expressed as:

$$\begin{aligned}\frac{g_{i,j,t+1} - g_{i,j}}{\Delta t} = & -\frac{g_{i,j,t} s_{i,j,l,m,b} \mathbf{1}_{s_{i,j,l,m,b}>0} - g_{i,j-1,t} s_{i,j-1,l,m,f} \mathbf{1}_{s_{i,j-1,l,m,f}>0}}{\Delta a} \\ & -\frac{g_{i,j+1,t} s_{i,j+1,l,m,b} \mathbf{1}_{s_{i,j+1,l,m,b}<0} - g_{i,j,t} s_{i,j,l,m,b} \mathbf{1}_{s_{i,j,l,m,b}<0}}{\Delta a} - \lambda_i g_{i,j,t} + \lambda_{-i} g_{-i,j,t}.\end{aligned}$$

In this case, let us define

$$\mathbf{g}_t = \begin{bmatrix} g_{1,1,t} \\ g_{1,2,t} \\ \vdots \\ g_{1,J,t} \\ g_{2,1,t} \\ \vdots \\ g_{2,J,t} \end{bmatrix},$$

as the density conditional on the current state of B_l and N_m. We assume that g_0 is the density in the DSS, and the update in the next time period is given by the KF equation:

$$\mathbf{g}_{t+1} = \left(\mathbf{I} - \Delta t \mathbf{A}_{l,m}^{\mathbf{T}}\right)^{-1} \mathbf{g}_t,$$

where $\mathbf{A}_{l,m}^{\mathbf{T}}$ is the transpose matrix of $\mathbf{A}_{l,m} = \lim_{n\to\infty} \mathbf{A}_{l,m}^n$, defined above.

Step 3: Update of the PLM using a neural network

The vector θ is recursively updated according to $\theta_{m+1} = \theta_m - \epsilon_m \nabla \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)$, where:

$$\nabla \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right) \equiv \left[\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_0^2}, \frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_1^2}, ..., \frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_{2,Q}^1}\right]^\top$$

is the gradient of the error function with respect to θ evaluated at $\left(\mathbf{s_j}, \widehat{h}_j\right)$.

The step size $\epsilon_m > 0$ is selected in each iteration according to a *line-search* algorithm in order to minimize the error function in the direction of the gradient. The algorithm is run until $\|\theta_{m+1} - \theta_m\| < \varepsilon$, for a small ε.

The error gradient can be efficiently evaluated using a *back-propagation* algorithm, originally developed by Rumelhart et al. (1986), which builds on the chain rule of differential calculus. In our case, the corresponding formulae are:

$$\begin{aligned}
\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_0^2} &= h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j \\
\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_q^2} &= \left(h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j\right) \phi\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s_j^i\right), \text{ for } q = 1, ..., Q \\
\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_{0,q}^1} &= \theta_q^2 \left(h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j\right) \phi'\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s_j^i\right), \text{ for } q = 1, ..., Q \\
\frac{\partial \mathcal{E}\left(\theta; \mathbf{s_j}, \widehat{h}_j\right)}{\partial \theta_{i,q}^1} &= s_j^i \theta_q^2 \left(h\left(\mathbf{s}_j; \theta\right) - \widehat{h}_j\right) \phi'\left(\theta_{0,q}^1 + \sum_{i=1}^{2} \theta_{i,q}^1 s_j^i\right), \text{ for } i = 1, 2 \text{ and } q = 1, ..., Q,
\end{aligned}$$

where $\phi'(x) = \frac{1}{(1+e^{-x})}$.

We fine-tune the training scheme of our neural network in several ways to meet the specific demands of the problem at hand.

First, the training scheme needs to yield a consistently good approximation: a "not-good-enough" approximation in any of the dozens of iterations of the algorithm can make it break and not deliver a result. This is the reason we employ a line-search instead of using a constant or adaptive learning rate: it prevents bad steps in the minimization.

Second, the training scheme cannot introduce big amounts of noise in the progressive approximation to the solution of the model. Otherwise, the noise can mask or prevent convergence (strict convergence criteria can only be met by chance, if at all). This is why we use a batch gradient descent (i.e., all training points are used in every gradient calculation and line-search step) instead of the more popular stochastic or mini-batch gradient descents. The random choice of training points in each step, common in the machine learning literature, would make progress stochastic and introduces noise.

Reducing stochastic elements in the solution method is also why we train the model with a grid approximation that clears out noise. We define a 101x101 grid over the (B, N) support, and assign each simulated point to one of the knots in that grid. Then, we run a linear regression, and use it to estimate the height of the PLM at that knot. This grid could later be used to solve the model using interpolation (e.g., with splines, or with natural neighbor interpolation) and linear extrapolation (and we do that as a robustness check, finding similar results to those

of the solution with the neural network). However, on a 2D surface. such extrapolation tends to generate ridges (because of the amplification of sample noise in far-away extrapolations), which could prevent convergence at the HJB step. Instead, we use those knots to train the neural network, which provides a good-enough fit in the visited area and a much smoother extrapolation to the nonvisited area of the (B, N) support.

Finally, the need to avoid stochastic elements that introduce noise in the algorithm, plus a desire for fast running times, made us change the usual Monte Carlo multi-start initialization of the neural network parameters for all iterations except the first one. In the first iteration of the algorithm, we do ten random initializations of the parameters of the neural network and ten subsequent training sessions. Later, we choose the best-performing trained network across those ten training sessions. From the second iteration onward, the neural network is initialized using weights that were found to be optimal in the previous iteration, and a single training session is carried out. This avoids re-introducing a stochastic element at each step. Using a small relaxation parameter in the PLM update step (it starts at 0.30 and exponentially decays toward 0.05, reaching 0.20 after five iterations and 0.10 after 16) makes the convergence of the full algorithm slower but smoother, with slow updates to the optimal neural network that help this nonrandom initialization work well.

Complete algorithm

We can now summarize the complete algorithm. We begin with a guess of the PLM $h^0(B, N)$. Set $s := 1$:

Step 1: Household problem. Given $h^{s-1}(B, N)$, solve the HJB equation to obtain an estimate of the value function $\mathbf{v}$ and of the matrix $\mathbf{A}$.

Step 2: Distribution. Given $\mathbf{A}$, simulate T periods of the economy using the KF equation and obtain the aggregated ebt $\{B_t\}_{t=0}^{T}$ and equity $\{N_t\}_{t=0}^{T}$. The law of motion of equity is

$$N_t = N_{t-1} + \left[\alpha Z (B_t + N_t)^{\alpha} - \delta (B_t + N_t) - r_t B_t - \widehat{\rho} N_t\right] \Delta t + \sigma (B_t + N_t) \sqrt{\Delta t} \varepsilon_t,$$

where $\varepsilon_t \overset{iid}{\sim} N(0, 1)$.

Step 3: PLM. Update the PLM using a neural netwok: h^s. If $\|h^s - h^{s-1}\| < \varepsilon$, where ε is a small positive constant, then stop. If not, return to step 1.

Note that, in order to avoid scale problems, we normalize all the inputs of the neural network so that the range is determined to be [-1,1]. To this end, we find the maximum and minimum values of the training sample and compute the mid-point between them. Then, we subtract the mid-point from every individual sample and divide the result by half of the max-min interval.

C Evaluating the likelihood function with micro observations

A promising avenue to improve the estimation in the main text is to add micro observations, which bring much additional information and help in integrating different levels of aggregation to assess the empirical validity of the model. More concretely, let $X_t \equiv [g_t(a,z); N_t]'$ be a vector of observations on the asset holdings of agents in this economy (households, $g_t(a,z)$, and the expert, N_t). Imagine, as before, that we have $D+1$ observations of X_t at fixed time intervals $[0, \Delta, 2\Delta, .., D\Delta,]$:

$$X_0^D = \{X_0, X_\Delta, X_{2\Delta}, ..., X_D\}.$$

At this moment, we need to assume –as is typically done in models with heterogeneous agents and aggregates shocks– that the *conditional no aggregate uncertainty* (CNAU) condition holds. See, for instance, Miao (2006), following Bergin and Bernhardt (1992). This condition implies that if households are distributed on the interval $I = [0,1]$ according to the Lebesgue measure Φ, then

$$G_t(A \times Z) = \Phi\left(i \in I : \left(a_t^i, z_t^i\right) \in A \times Z\right),$$

for any subsets $A \subset [0, \infty)$, $Z \subset \{z_1, z_2\}$. That is, the probability under the conditional distribution is the same as the probability according to the Lebesgue measure across I.

The likekihood that an individual agent $i \in I$ at time $t = d\Delta$ is at state $(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})$ is $f_{d\Delta}^d(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})$. The log-likelihood is then $\log\left[f_{d\Delta}^d(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})\right]$. Notice that this log-likelihood is a function of i.

The conditional aggregate log-likelihood across all agents is:

$$\log p_X\left(X_{d\Delta} | X_{(d-1)\Delta}; \Psi\right) = \int \log\left[f_{d\Delta}^d(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})\right] \Phi(di),$$

and, taking into account the CNAU condition, we get:

$$\begin{aligned}
\int \log\left[f_{d\Delta}^d(a_{d\Delta}^i, z_{d\Delta}^i, B_{d\Delta}, N_{d\Delta})\right] \Phi(di) &= \int \log\left[f_{d\Delta}^d(a, z, B_{d\Delta}, N_{d\Delta})\right] G_{d\Delta}(da, dz) \\
&= \sum_{i=1}^{2} \int_0^\infty \log\left[f_{d\Delta}^d(a, z_i, B_{d\Delta}, N_{d\Delta})\right] g_{d\Delta}(a, z) da,
\end{aligned}$$

where, in the second line, we have applied the definition of the Radon-Nikodym derivative to get the differential in a.

The density $f_t^d(a, z, B, N)$ follows the KF equation:

$$\begin{aligned}\frac{\partial f_t^d}{\partial t} = &-\frac{\partial}{\partial a}\left(s_t\left(a, z_i\right) f_t^d(a, z_i, B, N)\right) - \lambda_i f_t^d(a, z_i, B, N) + \lambda_j f_t^d(a, z_j, B, N) \\ &-\frac{\partial}{\partial B}\left[h(B, N) f_t^d(a, z_i, B, N))\right] - \frac{\partial}{\partial N}\left[\mu_t^N(B, N) f_t^d(a, z_i, B, N)\right] \\ &+\frac{1}{2}\frac{\partial^2}{\partial N^2}\left[\left(\sigma_t^N(B, N)\right)^2 f_t^d(B, N)\right], \quad (i \neq j = 1, 2) \end{aligned} \tag{48}$$

where

$$f_{(d-1)\Delta}^d = g_{(d-1)\Delta}(a, z)\delta\left(B - B_{(d-1)\Delta}\right)\delta\left(N - N_{(d-1)\Delta}\right),$$

which, following the same reasoning as in the previous subsection, is easy to evaluate.

More concretely, we use the notation $f_{i,j,l,m}^d \equiv f_i^d(a_j, B_l, N_m)$ and define a time step $\Delta t = \frac{\Delta}{S}$, where $1 << S \in \mathbb{N}$ is a constant. If we solve the KF equation (48) using a finite difference scheme, we have, for $t = (d-1)\Delta$ and $s = 1, .., S-1$,

$$\begin{aligned}\mathbf{f}_{t+s\Delta t}^d &= \left(\mathbf{I} - \Delta t \mathbf{A}^{\mathbf{T}}\right)^{-1} \mathbf{f}_{t+(s-1)\Delta t}^d, \\ \mathbf{f}_t^d &= \mathbf{g}_t \delta_{N_{(d-1)\Delta}} \delta_{B_{(d-1)\Delta}},\end{aligned}$$

where δ is the Kronecker delta and $\mathbf{f}_t^d$ is defined as

$$\mathbf{f}_t^d = \begin{bmatrix} f_{1,1,1,t} \\ f_{1,1,1,2,t} \\ \vdots \\ f_{2,J,L,M,t} \end{bmatrix}.$$

The conditional density $p_X\left(X_{d\Delta}|X_{(d-1)\Delta};\gamma\right)$ can be approximated by:

$$p_X\left(X_{d\Delta}|X_{(d-1)\Delta};\gamma\right) = \sum_{i=1}^{2}\sum_{j=1}^{J}\sum_{l=1}^{L} f_{i,j,l^d,m}^d g_{i,j}^d \Delta a \Delta B,$$

where $f_{i,j,l^d,m}^d$ is the density evaluated at the observed equity point $N_{d\Delta}$, $f_i^d(a_j, B_l, N = N_{d\Delta})$ and $g_{i,j}^d$ are the elements of the observed distribution $\mathbf{g}_{d\Delta}$.

D Chebyshev polynomials

Now, we show that the approximation to the PLM computed with Chebyshev polynomials is not satisfactory (even if we forget the considerations about the curse of dimensionality and coding complexity that we highlighted in the main text). In Figure D.1, we plot the PLM obtained with an algorithm similar to ours, but where we substitute a linear combination of Chebyshev polynomials for the neural network and we select the coefficients of that linear combination to fit the simulated data as well as possible.

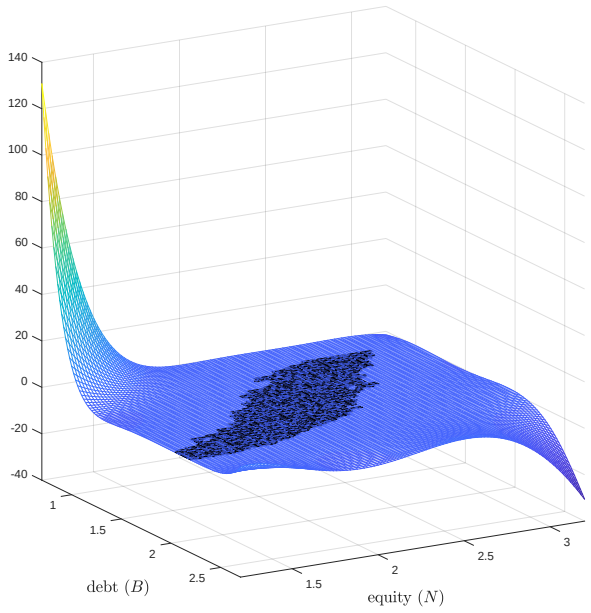

Figure D.1: PLM with Chebyshev polynomials.

While, at first inspection, the PLM seems sensible, a closer examination of the scale of the y-axis in Figure D.1 reveals large and implausible movements in $h(B, N)$. These variations can be seen better in Figure D.2, where we zoom $h(B, N)$ on a smaller range of debt and equity. The PLM is well approximated in the ergodic distribution (shaded area in the center) but, as soon as we move slightly outside that area, the oscillating features of polynomial approximations reassert themselves. Using this PLM means, in practice, unstable simulations and unreliable results.

Similar problems appear in solutions construed with splines or other popular series approximations: extrapolation requires a well-behaved basis and neural networks do an excellent job at such a task.

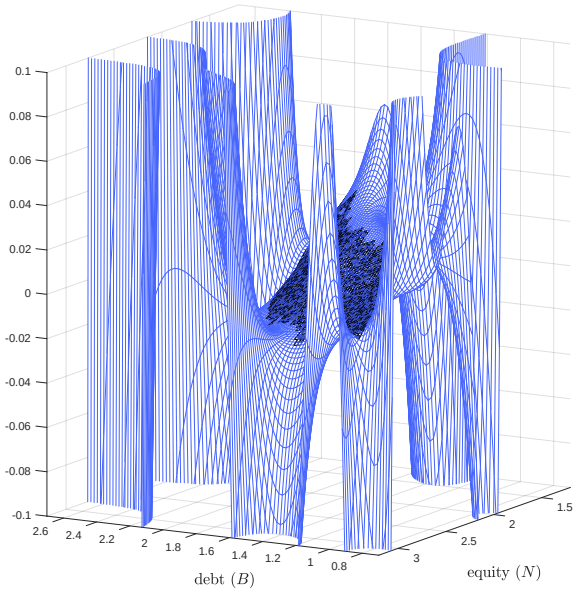

Figure D.2: PLM with Chebyshev polynomials (zoom).

E The values of the SSS(s)

In this section, we explore how the values of the SSS(s) change as we move some parameter values of the model. The exercises clarify some of the arguments in the main text regarding the supply and demand of debt by the expert and the households.

Figure E.3 plots the values of the LL-SSS, the unstable SSS, and the HL-SSS (plus, for reference, the DSS) as a function of σ. We can see how the leverage in the HL-SSS is a negative function of σ, a roughly constant function in the unstable SSS, and an increasing function in the LL-SSS (until the additional SSS(s) disappear). The mechanism for these three slopes is the same as the one discussed in the main text. In the HL-SSS, as σ grows, the expert wants to unload some of the capital risk by reducing its leverage. In comparison, in the LL-SSS, the households demand more debt as σ increases. This figure complements Figure 11 in the main text.

Figure E.4 repeats the same exercise, but now plotting the levels of leverage for the LL-SSS, the unstable SSS, the HL-SSS, and the DSS as a function of z_1. All levels are falling as z_1 increases, until we get to very little debt, a consequence, for high z_1, of small precautionary saving and the higher discount factor of the households. This figure complements Figure 14 in the main text.

Next, Figure E.5 plots the values of the SSS(s) as we simultaneously move aggregate and idiosyncratic risk. This figure complements Figure 17 in the main text.

Finally, Figure E.6 shows the value of the SSS(s) as we vary $\widehat{\rho}$, the discount factor of the expert. As the expert becomes more impatient, the level of leverage in the HL-SSS and LL-SSS

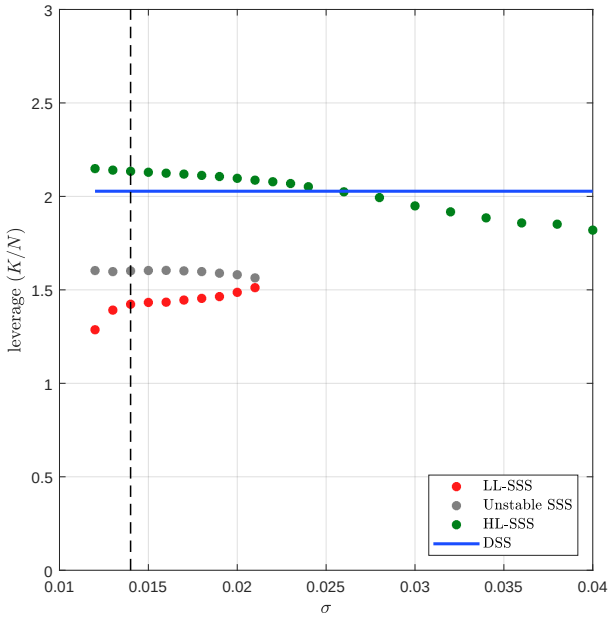

Figure E.3: SSS(S) as a function of σ.

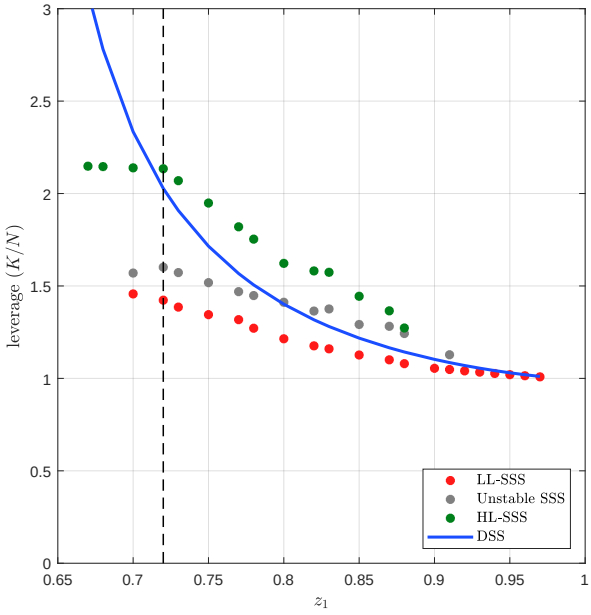

Figure E.4: SSS(s) as a function of z_1.

increases slightly, while the level of leverage at the DSS rises much more strongly. The reason is that as $\widehat{\rho}$ grows, households are more patient and, therefore, more willing to accumulate bonds and increase leverage.

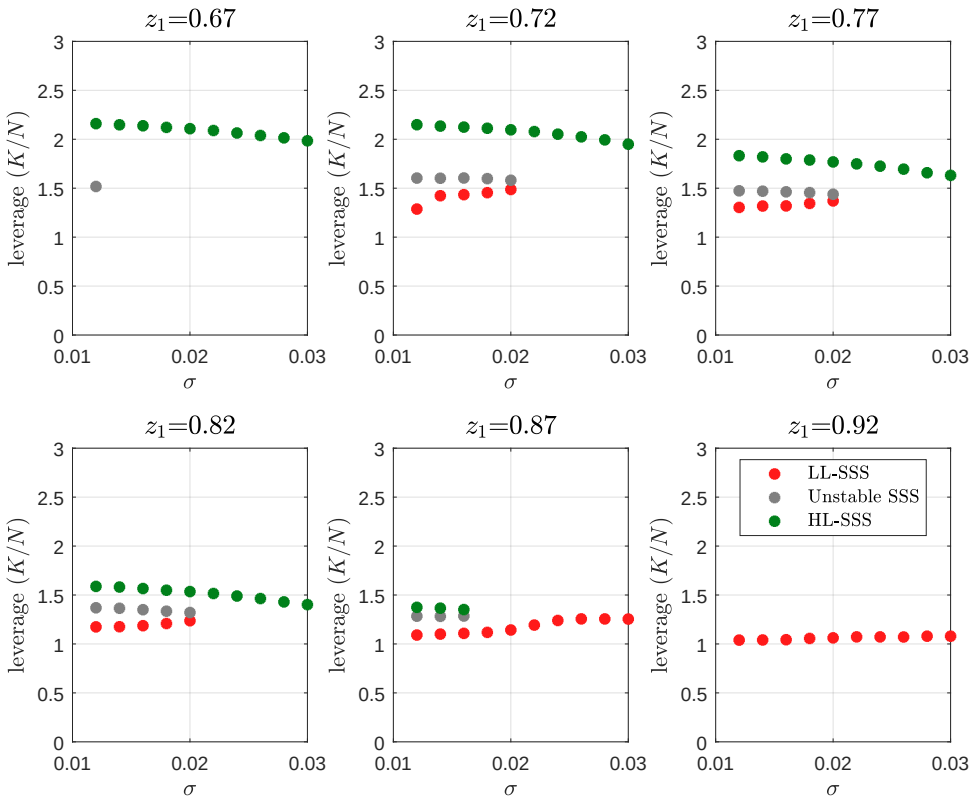

Figure E.5: SSS(s) as a function of σ and z_1.

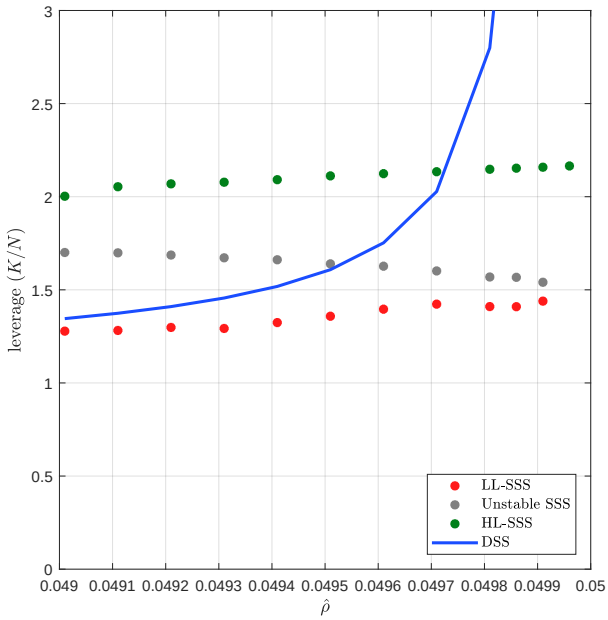

Figure E.6: SSS(s) as a function of $\widehat{\rho}$.

F Value functions

The second row of Figure F.7 plots the value functions of the households as a function of assets for low- and high-labor productivity at the HL-SSS and the LL-SSS. For easy reference,

in the first row of Figure F.7, we reproduce the distributions of households.

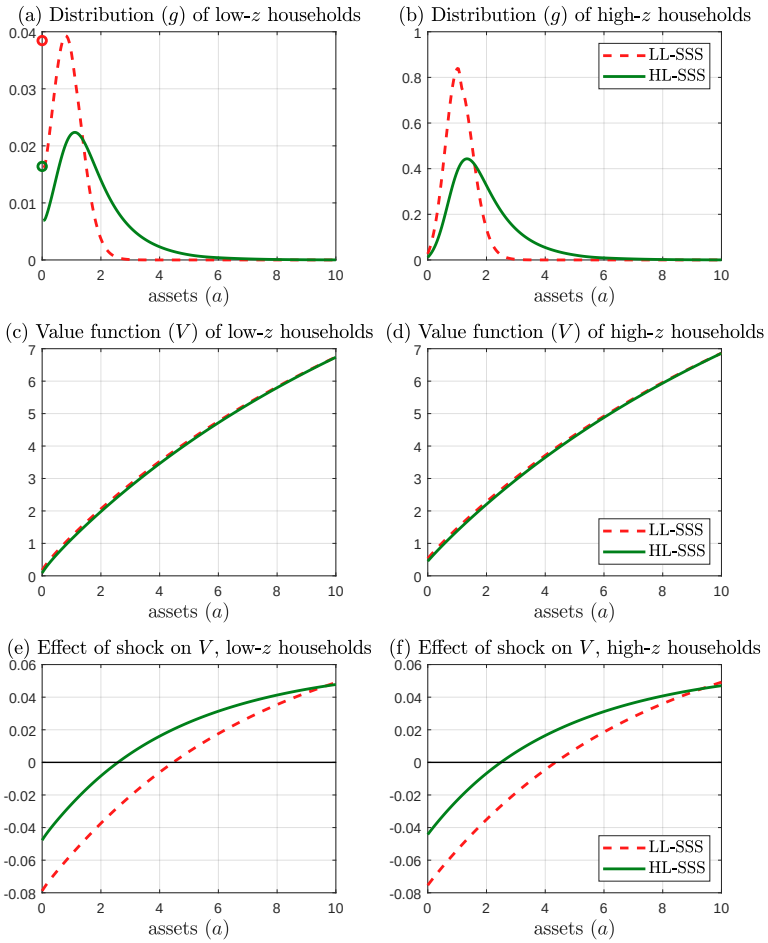

Figure F.7: Wealth distribution and value functions in the DSS and SSS.

The comparison of value functions shows that, for all levels of assets, households prefer to be at the LL-SSS than at the HL-SSS. This fact is not a surprise since at the LL-SSS the economy is less volatile and households have concave preferences only over consumption (not allowing, therefore, substitution with leisure when productivity is low). However, precautionary behavior also means that, at the HL-SSS, we will have more rich households.

Lastly, the bottom row of Figure F.7 shows how the value function changes after a two-standard-deviations negative capital shock: poorer households are worse off (they have lower wages), but wealthier households are better off, as their bonds pay a higher interest rate. The effect is more acute at the HL-SSS, as the persistence of wages and the risk-free interest rate are higher.